

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number  
WO 01/88114 A2

- (51) International Patent Classification<sup>7</sup>: C12N 15/00 LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (21) International Application Number: PCT/EP01/05794
- (22) International Filing Date: 18 May 2001 (18.05.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 0012231.7 19 May 2000 (19.05.2000) GB
- (71) Applicant (for all designated States except US): DEV-GEN NV [BE/BE]; Technologiepark 9, B-9052 Zwijnaarde (BE).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): PLAETINCK, Geert [BE/BE]; Pontstraat 16, B-9820 Merelbeke (BE). RENARD, Jean-Pierre [BE/BE]; Peter Benoîtlaan 141, B-9050 Gentbrugge (BE). MORTIER, Katherine [BE/BE]; Lange Boomgaardstraat 25, B- Gent (BE). BOGAERT, Thierry [BE/BE]; Wolvendreef 26g, B-8500 Kortrijk (BE).
- (74) Agent: BOULT WADE TENNANT; Verulam Gardens, 70 Gray's Inn Road, London WC1X 8BT (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Declarations under Rule 4.17:**
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG)
- of inventorship (Rule 4.17(iv)) for US only
- Published:**
- without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/88114 A2

(54) Title: GENE EXPRESSION SYSTEM

(57) Abstract: There are provided DNA constructs, including replicable cloning vectors and expression vectors, comprising a bacteriophage promoter operably linked to an outon sequence. The expression vectors provided by the invention are useful in the expression of recombinant polypeptides in host cells or organisms and are particularly useful in expression of recombinant polypeptides in nematode worms such as *C. elegans*.

NUC 38884

- 1 -

## GENE EXPRESSION SYSTEM

## Field of the invention

5           The invention relates to the expression of DNA, genes, cDNAs, proteins, peptides and parts thereof in the nematode worm *C. elegans*. In particular, the invention relates to methods of improving the translation of RNAs transcribed in *C. elegans* using a  
10 bacteriophage polymerase by introduction of a trans-splice recognition site recognised by an SL1 trans-splice recognition sequence into the DNA template transcribed by the bacteriophage polymerase.

## 15   Background to the invention

Eukaryotic versus prokaryotic expression.

          Bacteriophage RNA polymerases, such as T7, T3, and SP6, and their corresponding promoters have been  
20 used extensively to drive the expression of heterologous genes in a variety of organisms. In co-pending International patent application No. WO 00/01846, Plaetinck et al. describe the use of the T7 system to express DNA, genes, cDNA, proteins and  
25 peptides of parts thereof and for the expression of double-stranded RNA (dsRNA) in the nematode model system *C. elegans*.

          The bacteriophage expression systems are well known in the art for use in prokaryotic host cells,  
30 such as *E. coli*, and have the advantage that they provide simple and strong expression systems dependent only on one RNA polymerase and one well defined promoter. The application of such efficient expression systems in eukaryotic organisms is,  
35 however, not evident, mainly because messenger RNAs from eukaryotes and prokaryotes have a different

- 2 -

structure, which has implications for translation efficiency and RNA stability.

1 Messenger RNAs of higher eukaryotes share a functionally essential 5' CAP structure. This  
5 structure is generated during a capping reaction that is linked exclusively to RNA polymerase II transcription. Prokaryotic RNA polymerases such as bacteriophage T3, T7 and SP6 polymerases do not  
10 provide messenger RNAs with such a CAP structure, leading to inefficient translation in eukaryotic systems (Fuerst et al. J. Mol. Biol:206:333-348 (1989)).

One way to improve translation of uncapped mRNAs in eukaryotic systems is by the insertion of an  
15 internal ribosome entry site (IRES) sequence 5' of the coding sequence. For example, Elroy-Stein, et al., Proc. Natl. Acad. Sci. USA 87:6743-6747 (1990), describe the cloning of the untranslated region of the ECMV virus downstream of the T7 promoter in order to  
20 enhance the efficiency of translation. In other systems translation of T7-derived transcripts may be enhanced by addition of a CAP structure derived from a capped transcript. For example, in Trypanosoma a 5' CAP structure is added to T7 generated RNA transcripts  
25 by a natural occurring trans-splicing reaction (Wirtz et al. NAR 22:3887-3894 (1994)).

#### Trans-splicing in *C. elegans*.

In *C. elegans* many mRNAs contain an identical  
30 short leader sequence, designated the spliced leader (SL). This splice leader is donated by a small RNA (SL RNA) via a trans-splicing reaction. This trans splicing was first observed by Krause et al., Cell 49:753-61 (1987). The splice leader RNA exists as a  
35 small nuclear ribonucleoprotein particle and has the trimethylguanosine cap that is characteristic of

- 3 -

eukaryotic small nuclear RNAs. The trimethylguanosine cap present on the spliced leader RNA is transferred to the pre-mRNA during the trans-splicing reaction. Thereafter, the trimethylguanosine cap is maintained on the mature mRNA (Van Doren et al., Mol. Cell. Biol. 10:1769-1772 (1990)). The trans-splicing signal for such a splice leader is essentially an intron missing only the 5' splice site, designated an 'outtron'. An outtron has essentially all the intron sequence including a trans-splice acceptor site homologous to a UUUCAG sequence preceded by a AU rich region (Conrad et al., NAR 21:913-919 (1993)). Introduction of an outtron into the 5' untranslated region of a *C. elegans* gene converts it to a trans-spliced gene (Conrad et al., EMBO J. 12:1249-1255 (1993); Conrad et al. Mol. Cell Biol. 11:1931-1926 (1991)) and introduction of donor sites in a natural trans-spliced *C. elegans* gene prevents trans-splicing and converts it into a more conventional gene.

20

#### Description of the invention.

Until recently, expression of heterologous and homologous genes in *C. elegans* was mainly achieved by linking an appropriate coding sequence to a selected *C. elegans* promoter. The present inventors have recently demonstrated that the recombinant gene expression in *C. elegans* can be based on the prokaryotic T7 expression system (WO 00/01846). However, the present inventors found that the expression system was far from being efficient, or at least the resulting expression was much lower than would be expected from this T7 related expression system. It was concluded that this low expression was mainly due to RNA instability or translation arrest. Furthermore, it was reasoned that fundamental differences between prokaryotic and eukaryotic

35

- 4 -

expression systems, particularly the requirement for capping of the 5' end of the mRNA for efficient translation in eukaryotic systems, was the main reason for this unexpectedly low expression.

5       The inventors have now developed a solution to the problem of the inefficiency of the T7 system in eukaryotic host cells and organisms, particularly in *C. elegans*, and have constructed a generally applicable expression system which allows for the  
10       efficient expression of genes, DNA, cDNA, peptides and proteins under the regulation of the T7 promoter in *C. elegans*.

          Therefore, in accordance with a first aspect of the invention there is provided a DNA construct  
15       comprising a bacteriophage promoter operably linked to an outtron sequence.

          It is an essential feature of the DNA construct of the invention that the bacteriophage promoter and the outtron sequence are "operably linked", that is to  
20       say they are arranged in a relationship permitting them to function in their intended manner. In this case, the bacteriophage promoter is positioned upstream of the outtron sequence such that it is capable of promoting transcription of the outtron  
25       sequence upon binding of an appropriate RNA polymerase, with the outtron sequence forming the extreme 5' end of the resulting transcript.

          The DNA construct may further comprise at least one restriction enzyme recognition site positioned  
30       downstream of and proximal to the outtron sequence. Advantageously, the DNA construct may contain multiple restriction sites forming a multi-cloning site. The purpose of the restriction site/multi-cloning site is to facilitate cloning of a heterologous or homologous  
35       DNA fragment downstream of the outtron sequence. A DNA construct comprising a bacteriophage promoter, an outtron sequence and a restriction site/multi-cloning

- 5 -

site may therefore be referred to hereinafter as an 'outtron cloning construct'.

! In an outtron cloning construct it is advantageous for the restriction site/multi-cloning site to be positioned fairly proximal to the outtron sequence (e.g. within 100bp) such that a heterologous or homologous sequence inserted at this site may be co-transcribed with the outtron sequence on a single mRNA. However, further sequence elements may be interposed between the outtron sequence and the restriction site/multi-cloning site. For example, the general purpose vector pDW3123 described in the accompanying examples has a synthetic intron A sequence between the outtron sequence and the multi-cloning site.

In one preferred embodiment of the invention, the DNA construct is a replicable cloning vector, such as, for example, a plasmid vector. In addition to the bacteriophage promoter, outtron sequence and optional restriction site/multi-cloning site, the vector may further contain one or more of the general features commonly found in cloning vectors, for example an origin of replication to allow autonomous replication within a host cell and a selective marker, such as an antibiotic resistance gene. Although not essential, the vector may also contain a poly-adenylation signal to stabilize and process the 3' end of the mRNA transcribed from the bacteriophage promoter. A preferred example is the 3'UTR from the *C. elegans* unc-54 gene, but any other 3'UTR or polyadenylation signal may be used.

Outtron-containing DNA constructs according to the invention may be easily be constructed from the component sequence elements using standard recombinant techniques well known in the art and described, for example, in F. M. Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1994).

- 6 -

Outron sequences for use in the constructs of the invention may be isolated from natural *C. elegans* genes using standard molecular biology techniques. For example, a natural outron sequence might be  
5 amplified using the polymerase chain reaction or an equivalent amplification technique using *C. elegans* genomic DNA as a template. Alternatively, synthetic outron sequences may be synthesised, for example, by annealing two complementary single  
10 stranded oligonucleotides, as illustrated in the accompanying examples. Once a DNA fragment comprising the outron sequence has been obtained it would be a matter of routine to assemble an outron construct by linking the outron in the correct orientation relative  
15 to the bacteriophage promoter.

The sequences of the commonly used bacteriophage promoters, e.g. T7, T3 and SP6, are well known in the art and oligonucleotides containing functional phage promoter sequences can be readily synthesised using  
20 standard oligonucleotide synthesis techniques. It would be a matter of routine to insert such a synthetic promoter sequence into, for example, a plasmid vector backbone containing, for example, an origin of replication a selective marker and a  
25 suitable restriction site. Alternatively, one of the many plasmid vectors containing bacteriophage promoter sequences known in the art may be used as the starting point for the construction of a plasmid-based outron cloning vector. The known vectors generally contain,  
30 in addition to the phage promoter sequence, one or more restriction sites conveniently positioned downstream of the phage promoter and also a bacterial origin of replication and a selective marker. Once the vector backbone is in place the outron sequence  
35 may simply be inserted in the appropriate position downstream of the bacteriophage promoter.

In a particularly useful embodiment the invention

- 7 -

provides a DNA construct for use in bacteriophage promoter-driven expression of a polypeptide in a eukaryotic host cell or organism. This construct comprises a bacteriophage promoter operably linked to  
5 a DNA sequence such that it is capable of initiating transcription of the DNA sequence upon binding of an appropriate RNA polymerase to the promoter, wherein the aforesaid DNA sequence comprises an outtron sequence and at least one open reading frame  
10 positioned downstream of the outtron sequence.

The open reading frame may be essentially any protein-encoding DNA sequence bounded by start and stop codons. This protein-encoding DNA sequence may include introns, as both trans-splicing and cis-  
15 splicing can occur together.

A DNA construct according to this embodiment of the invention, which may be referred to hereinafter as an 'outtron expression construct', may be derived from an outtron cloning construct by insertion of a  
20 heterologous or homologous protein-encoding DNA fragment into the restriction site/multi-cloning site. It is essential that the heterologous or homologous DNA fragment be inserted downstream of the outtron sequence such that the two sequences may be co-  
25 transcribed, with the outtron sequence forming part of the 5' untranslated region of the resulting mRNA.

The outtron expression construct may advantageously form an expression vector, such as, for example, a plasmid vector. Most preferably, the  
30 expression vector will be one suitable for use in the nematode worm *C. elegans*. In addition to the bacteriophage promoter, outtron sequence and protein-encoding DNA sequence (open reading frame), the expression vector may further contain one or more of  
35 the general features commonly found in expression vectors, for example an origin of replication to allow autonomous replication within a bacterial host cell



- 8 -

and a selective marker, such as an antibiotic resistance gene. The vector may also contain a poly-adenylation signal to stabilize and process the 3' end of the mRNA transcribed from the bacteriophage promoter. A preferred example is the 3'UTR from the *C. elegans* unc-54 gene, but any other 3'UTR or polyadenylation signal may be used. An additional element, such as for example a synthetic intron, may be interposed between the outtron sequence and the open reading frame.

It is important that the open reading frame is positioned downstream of and proximal to the outtron sequence in the expression construct such that (i) the two elements are co-transcribed to form a single mRNA and (ii) the outtron sequence forms part of the 5' untranslated region of the mRNA. If the appropriate splicing machinery and a supply of SL RNAs is provided by the eukaryotic host cell or organism then the uncapped 5' end of the pre-mRNA transcribed from the expression construct will be replaced with a capped splice leader via the trans-splicing reaction. This will greatly increase the efficiency of translation in a eukaryotic host system.

The use of an outtron sequence at the extreme 5' end of the RNA provides a solution to the problem of reduced expression efficiency in eukaryotic systems wherever the type of promoter/polymerase used to drive gene expression leads to the production of uncapped transcripts, provided that the host cell or organism produces the spliced leader RNAs required for the trans-splicing reaction.

Outtron sequences which may be utilised in accordance with the invention include naturally occurring outtron sequences isolated from SL1-specific *C. elegans* genes (Conrad, R. Functional analysis of a *C. elegans* trans-splice acceptor. *Nucleic Acids Res.* 1993, 21(4), pp913-919; Conrad, R. SL1 trans-splicing

- 9 -

specified by AU-rich synthetic RNA inserted at the 5' end of *Caenorhabditis elegans* pre-mRNA. RNA. 1995, 1(2), pp164-170) and also synthetic outtron sequences which are functionally equivalent to the natural *C. elegans* outtron sequences, including variants of naturally occurring *C. elegans* outtrons. The phrase "functionally equivalent" means that the synthetic intron is recognised by the *C. elegans* trans-splicing machinery and can be trans-spliced to a *C. elegans* splice leader RNA, preferably the SL1 splice leader.

Experimental evidence indicates that trans-splicing in *C. elegans* is signalled by an AU-rich intron-like sequence followed by a splice acceptor site (Conrad *et al* 1993 and 1995). For the purposes of the present application the terms "outtron" or "outtron sequence" should be interpreted as referring to both the AU-rich region from the 5' end of the pre-mRNA to the trans-splice acceptor site and the trans-splice acceptor site itself. In connection with the DNA constructs of the invention, the terms "outtron" and "outtron sequence" refer to features present in the DNA which encodes the pre-mRNA.

The consensus splice acceptor site for trans-splicing of outtrons and the consensus 3' splice acceptor site for cis-splicing of introns are essentially identical (UUUCAG). Moreover, a normally trans-spliced acceptor site can be efficiently cis-spliced when a donor splice site is inserted upstream within the outtron sequence. It is therefore important that the outtron constructs described herein do not contain any potential splice donor sequence upstream of the splice acceptor within the outtron and downstream of the transcription start site such that it will be transcribed in the mRNA encoded by the construct. If such a site were present then there would be a potential for cis-splicing rather than

- 10 -

trans-splicing.

It has also been observed that the overall length of the outtron has an effect on the efficiency of trans-splicing, longer outtrons in general working better than shorter ones (Conrad *et al.* 1995). Advantageously, the outtron sequences for inclusion into the outtron constructs described herein should be greater than about 50nt in length.

A synthetic outtron containing an AT stretch and a TTTTCAG sequence has been shown to be functional in *C. elegans*. As illustrated in the accompanying Examples, the insertion of an outtron sequence into the 5' untranslated region of GFP reporter construct, downstream of the promoter and upstream of the GFP open reading frame, is required for optimal expression of bacteriophage RNA polymerase transcribed reporter gene mRNA in *C. elegans*.

Suitable bacteriophage promoters which may be used in the DNA constructs according to the invention include T7, T3 and SP6 promoters, with T7 being the most preferred. As discussed above, these bacteriophage promoters have long been known to be useful tools in molecular biology since they can provide simple and strong expression systems dependent only on the binding of the specific or cognate RNA polymerase.

In a still further aspect, the invention provides a method for expressing a recombinant polypeptide in *C. elegans*, which method comprises:

introducing an outtron expression construct, as described above, said construct being an expression vector suitable for use in *C. elegans*, into a *C. elegans* strain which expresses an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell

- 11 -

types.

An outtron expression vector for use in this method may be constructed by inserting DNA encoding the polypeptide of interest into an outtron cloning  
5 vector, as described above. The vector must be one which is suitable for use in *C. elegans*, plasmid-based vectors are the most preferred.

The *C. elegans* worms are preferably transgenic worms carrying a transgene capable of expressing the  
10 RNA polymerase in one or more tissues or cell types. The term "transgene capable of expressing" as used herein means a nucleic acid molecule comprising a nucleotide sequence encoding the polymerase operably linked to a promoter. The promoter may be any  
15 promoter which functions in *C. elegans* and may be general (i.e. active in substantially all tissues and cell types), tissue-specific, cell type-specific, constitutive, inducible etc. Most preferably, the promoter will exhibit tissue or cell type-specificity.  
20 With the use of a tissue or cell type-specific promoter of the appropriate specificity it is possible to control the site of RNA polymerase expression within *C. elegans* and hence control the site of expression of the recombinant polypeptide.

25 Methods for the construction of transgenic *C. elegans* worms are known in the art and are particularly described by Craig Mello and Andrew Fire, Methods in Cell Biology, Vol 48, Ed. H.F. Epstein and D.C. Shakes, Academic Press, pages 452-480.

30 In a further aspect the invention provides a kit for use in recombinant expression of a polypeptide in *C. elegans*, the kit comprising an outtron cloning construct, as described above, and optionally a supply  
35 of *C. elegans* nematode worms expressing an RNA polymerase specific for the bacteriophage promoter

- 12 -

present in the said outtron cloning construct in one or more tissues or cell types.

1     The kit might further contain control inserts and control constructs, e.g. a reporter gene inserts and  
5     constructs which could be used to check efficiency of cloning steps and transfection steps, respectively. It might also contain constructs which may be used as selectable markers in the transfection procedure, e.g. a rol 6 plasmid (see below).

10     The invention further provides methods for the construction of transgenic *C. elegans* expressing a recombinant polypeptide in one or more tissues or cell types. One such method comprises introducing an  
15     outtron expression construct, as described above, said construct being an expression vector suitable for use in *C. elegans* comprising an open reading frame encoding the desired recombinant polypeptide, into a  
20     *C. elegans* strain which expresses an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell  
25     types, and isolating transgenic *C. elegans* lines which stably express the said polypeptide. The *C. elegans* strain expressing the polymerase is preferably a transgenic strain carrying a transgene capable of  
30     expressing the RNA polymerase in one or more tissues or cell types, as described above. As aforesaid, transgenic *C. elegans* lines can readily be constructed using standard techniques well known in the art.

35     In an alternative approach, the method may comprise introducing into a background *C. elegans* strain (i) an outtron expression construct, as described above, said construct being an expression vector suitable for use in *C. elegans* comprising an open frame encoding the desired recombinant  
40     polypeptide, and (ii) a DNA construct suitable for expression of an RNA polymerase specific for the

- 13 -

bacteriophage promoter present in the outtron expression construct in one or more tissues or cell types of *C. elegans*, and isolating transgenic *C. elegans* lines which stably express the said polypeptide. The second DNA construct may, advantageously, be an expression vector comprising a nucleotide sequence encoding the polymerase operably linked to a promoter having the appropriate tissue or cell type specificity.

10 In carrying out the methods of the invention one may employ standard techniques well known in the art for construction and selection of transgenic *C. elegans* lines. Such techniques are described, for example, in techniques described in Methods in Cell Biology, vol 84; *Caenorhabditis elegans*: modern biological analysis of an organism, ed. Epstein and Shakes, academic press, 1995. Foreign DNA (e.g. plasmid DNA) may be introduced into *C. elegans* using microinjection or ballistic transformation, as described in the applicant's co-pending International patent application No. WO 99/49066. In order to facilitate the selection of transgenic strains a marker plasmid may be co-introduced with the transgenes. A typical example is the plasmid pRF4 (Mello, C. C. et al. EMBO J. 10, 3959-3970 (1991)) which carries the rol-6 gene. *C. elegans* expressing rol-6 can be identified by screening for the roller phenotype. Any other *C. elegans* dominant selectable phenotypic marker, of which there are many known in the art, may be used to facilitate selection of transgenic lines. A useful example is green fluorescent protein (or any of the equivalent autonomous fluorescent proteins known in the art).

35 In a still further aspect the invention provides transgenic *C. elegans* worms which contain an outtron expression construct, as described above, said

- 14 -

construct being an expression vector suitable for use  
in *C. elegans*, and which further express an RNA  
polymerase specific for the bacteriophage promoter  
present in the outtron expression construct in one or  
5 more tissues or cell types.

The present invention will be further understood  
with reference to the following non-limiting Examples,  
together with the accompanying drawings in which:

10

Figure 1 illustrates the construction of a T7-outtron-  
GFP vector. (A) sequence of the synthetic outtron  
produced by annealing oligonucleotides o-GN59 and o-  
GN60. (B) summary of the strategy used to construct  
15 vector pDW3124.

Figure 2 shows plasmid maps for pDW3123 (outtron  
cloning vector) and pDW3124 (outtron expression vector  
for GFP expression).

20

Figure 3 is a plasmid map of pGN148 which contains a  
T7 RNA polymerase coding sequence under the regulation  
of the *C. elegans* SERCA promoter.

25

Figure 4 illustrates the nucleotide sequence of  
pGN148.

Figure 5 illustrates the nucleotide sequence of pDW  
3123 annotated to show the positions of the T7  
30 promoter, outtron, synthetic intron A, multi-cloning  
site and unc-54 3' UTR sequences and also the  
ampicillin resistance gene.

35

Figure 6 illustrates the nucleotide sequence of pDW  
3124 annotated to show the positions of the T7  
promoter, outtron, synthetic intron A, GFP with introns  
and unc-54 3' UTR sequences and also the ampicillin

- 15 -

resistance gene.

Example 1 -Construction of a T7-outtron-GFP containing vector (pDW3124)

5       A SL1 trans-splice acceptor site (outtron) was cloned into a vector downstream of the T7 promoter and upstream of the GFP to be expressed.

10       A synthetic outtron consisting of two partially overlapping oligonucleotides (o-GN59 and o-GN60, see Figure 1) was inserted into a XbaI/XmaI digested T7 promoter GFP construct. Briefly, 25µl o-GN59 and 25µl o-GN60 (100µM) were denatured for 5 minutes at 94°C, annealed for 30 minutes at 68°C then cooled to 4°C. 1µl of XmaI/XbaI digested pDW3120 and 10µl of the  
15       annealed oligos were then ligated using T4 ligase overnight at 16°C, transformed into competent *E. coli* and analysed by restriction digestion and DNA sequencing, all according to standard molecular biology procedures. The resulting vector was  
20       designated pDW3124 (Figures 1 and 2).

      The outtron contains an AU rich sequence followed by a splice-acceptor site as described by Conrad et al, NAR 21:913-919 (1993) (see Figure 1).

25       Example 2-Construction of a T7-Outtron MCS vector

      A general purpose vector was constructed to facilitate expression of other DNA sequences in *C. elegans* under the control of the T7 promoter. This was done by digesting vector pDW3124 with HindII  
30       (position 179) and PvuII (position 1029) (partial digest) and re-ligating the blunt ends, resulting in vector pDW3123 (Figure 2).



- 16 -

Example 3-The expression of heterologous genes in *C. elegans* regulated by the T7 promoter requires trans-splicing.

5 Wild-type *C. elegans* nematodes were co-injected with various combinations of the following test plasmids:

- 1) GFP reporter plasmid  
GFP: pDW2020  
10 outtron-GFP: pDW2024  
T7 promoter-GFP: pDW3120  
T7 promoter-outtron-GFP: pDW3124
- 2) T7 polymerase expression plasmid SERCA T7  
15 polymerase: pGN148 together with pRF-4 (rol-6) as marker.

For every co-injection experiment, a total concentration of 200 ng DNA/ $\mu$ l was used (plasmid  
20 concentration was 50 ng/ $\mu$ l and carrier DNA was added up to 200ng/ $\mu$ l). For every co-injection  $\pm$ 15 adult worms were injected.

F1 offspring showing the marker rol-6 phenotype  
25 were isolated and then selected for further study. The next generation (F2) of the roller lines were screened for GFP expression in the pharynx, vulva, tail and body wall muscles. These are the tissues in which the bacteriophage T7 RNA polymerase is known to  
30 be expressed when under the control of the *C. elegans* SERCA promoter (as in the construct pGN148)

The results are shown in Table 1 below, which indicates the number of lines expressing GFP vs total number of lines isolated.

35

- 17 -

	1	2	3
A	Construct	no T7-polymerase construct	with T7-polymerase construct (50ng) pGN148
B	GFP (50ng) pDW2020	0/8	2/6*
C	outtron::GFP (50ng) pDW2024	0/11	3/8*
5 D	T7-promoter::GFP (50ng) pDW3120	0/3	0/5
E	T7-promoter::outtron::GFP (50ng) pDW3124	0/7	13/13

\* GFP-expression most probably result of recombination  
10 in the extrachromosomal array

No GFP expression was observed in the experiments  
where the T7 RNA polymerase was absent (cells B2, C2,  
15 D2, E2).

In the experiments where the T7 RNA polymerase  
expressing vector was co-injected with GFP vectors  
without a T7 promoter, as in the cells B3 and C3, GFP  
expression was sometimes observed. This is probably  
20 due to recombination events in the extrachromosomal  
arrays, resulting in transcription of GFP directly  
from the SERCA promoter.

In the experiments where the T7 promoter-GFP construct  
25 and the SERCA T7 RNA polymerase were co-injected, no  
GFP expression could be observed (cell D3). In  
contrast, all of the lines isolated from the  
experiments where the GFP transcript contained an  
outtron at its 5' site (n=13) expressed GFP (cell E3).  
30 The outtron is a favourable target for SL1  
trans-splicing. Since SL1 RNA molecules contain a 5'

- 18 -

trimethylguanosine CAP structure which is transferred to the mature mRNA this results in improved translation of the RNA and hence better expression of GFP. Without the outtron the T7 RNA polymerase transcripts do not carry a CAP structure at their 5' end, leading to inefficient translation. The results of this experiment illustrate the importance of trans-splicing for efficient expression of heterologous and homologous genes transcribed by prokaryotic polymerases in *C. elegans*.

SEQUENCE LISTING

SEQ ID NO: 1 Oligonucleotide o-GN59  
15 SEQ ID NO: 2 Oligonucleotide 0-GN60  
SEQ ID NO: 3 Plasmid pDW3123  
SEQ ID NO: 4 Plasmid pDW3124  
SEQ ID NO: 5 Plasmid pGN148

Claims:

1. A DNA construct comprising a bacteriophage promoter operably linked to an outtron sequence.

5

2. A DNA construct as claimed in claim 1 which further comprises at least one restriction enzyme recognition site positioned downstream of and proximal to the outtron sequence.

10

3. A DNA construct as claimed in claim 2 which comprises a multi-cloning site positioned downstream of and proximal to the outtron sequence.

15

4. A DNA construct as claimed in claim 2 or claim 3 which further comprises a DNA fragment inserted at the said restriction site or at a restriction site within the said multi-cloning site.

20

5. A DNA construct as claimed in any one of claims 1 to 4 which is a replicable cloning vector.

6. A DNA construct as claimed in any one of claims 1 to 5 wherein the outtron sequence comprises a 3' splice acceptor site having the sequence TTTCAG preceded by an AT-rich region.

25

7. A DNA construct as claimed in claim 6 wherein the outtron sequence comprises the nucleotide sequence illustrated in Figure 1A.

30

8. A DNA construct as claimed in any one of claims 1 to 7 wherein the bacteriophage promoter is the T7, T3 or SP6 promoter.

35

9. A DNA construct for use in bacteriophage promoter-driven expression of a polypeptide in a

1

- 20 -

eukaryotic host cell or organism, which construct comprises a bacteriophage promoter operably linked to a DNA sequence such that it is capable of initiating transcription of said DNA sequence upon binding of the appropriate RNA polymerase to the promoter, wherein the said DNA sequence comprises an outtron sequence and at least one open reading frame positioned downstream of the outtron sequence.

10. A DNA construct as claimed in claim 9 which is an expression vector.

11. A DNA construct as claimed in claim 9 or claim 10 wherein the outtron sequence comprises a 3' splice acceptor site having the sequence TTTCAG preceded by an AT-rich region.

12. A DNA construct as claimed in claim 11 wherein the outtron sequence comprises the nucleotide sequence illustrated in Figure 1A.

13. A DNA construct as claimed in any one of claims 9 to 12 wherein the bacteriophage promoter is the T7, T3 or SP6 promoter.

14. A kit for use in recombinant expression of a polypeptide in *C. elegans*, the kit comprising a DNA construct as claimed in any one of claims 1 to 3, and optionally *C. elegans* worms expressing an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell types.

15. A method for expressing a recombinant polypeptide in *C. elegans* which method comprises: introducing a DNA construct as claimed in any one

- 21 -

of claims 9 to 13, said construct being an expression vector suitable for use in *C. elegans*, into a *C. elegans* strain which expresses an RNA polymerase specific for the bacteriophage promoter present in  
5 said DNA construct in one or more tissues or cell types.

16. A method of generating transgenic *C. elegans* expressing a recombinant polypeptide, which method  
10 comprises:

introducing a DNA construct as claimed in any one of claims 9 to 13 comprising an open reading frame encoding the recombinant polypeptide, said construct being an expression vector suitable for use in *C. elegans*, into a *C. elegans* strain which expresses an  
15 RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell types, and

isolating transgenic *C. elegans* lines which  
20 stably express the said polypeptide.

17. A method of generating transgenic *C. elegans* expressing a recombinant polypeptide, which method comprises:

25 introducing into *C. elegans* (i) a first DNA construct as claimed in any one of claims 9 to 13 comprising an open reading frame encoding the recombinant polypeptide, said construct being an expression vector suitable for use in *C. elegans*, and  
30 (ii) a second DNA construct suitable for expression of an RNA polymerase specific for the bacteriophage promoter present in the first DNA construct in one or more tissues or cell types of *C. elegans*, and

isolating transgenic *C. elegans* lines which  
35 stably express the said polypeptide

- 22 -

18. Transgenic *C. elegans* which contain a DNA construct as claimed in any one of claims 9 to 13, said construct being an expression vector suitable for use in *C. elegans*, and which further express an RNA polymerase specific for the bacteriophage promoter present in said DNA construct in one or more tissues or cell types.

FIG. 1(A).

XbaI overhang	SspI	3' splice acceptor	
CTAGATTACAAC TAATTATACTTATTTGAATATTCAAATTTTCAGAC			o-GN59
TAATGTTGATTAAATATGAATAAACTTATAAGTTTAAAAGTCTGGGCC			o-GN60
		XmaI overhang	

FIG. 1(B).

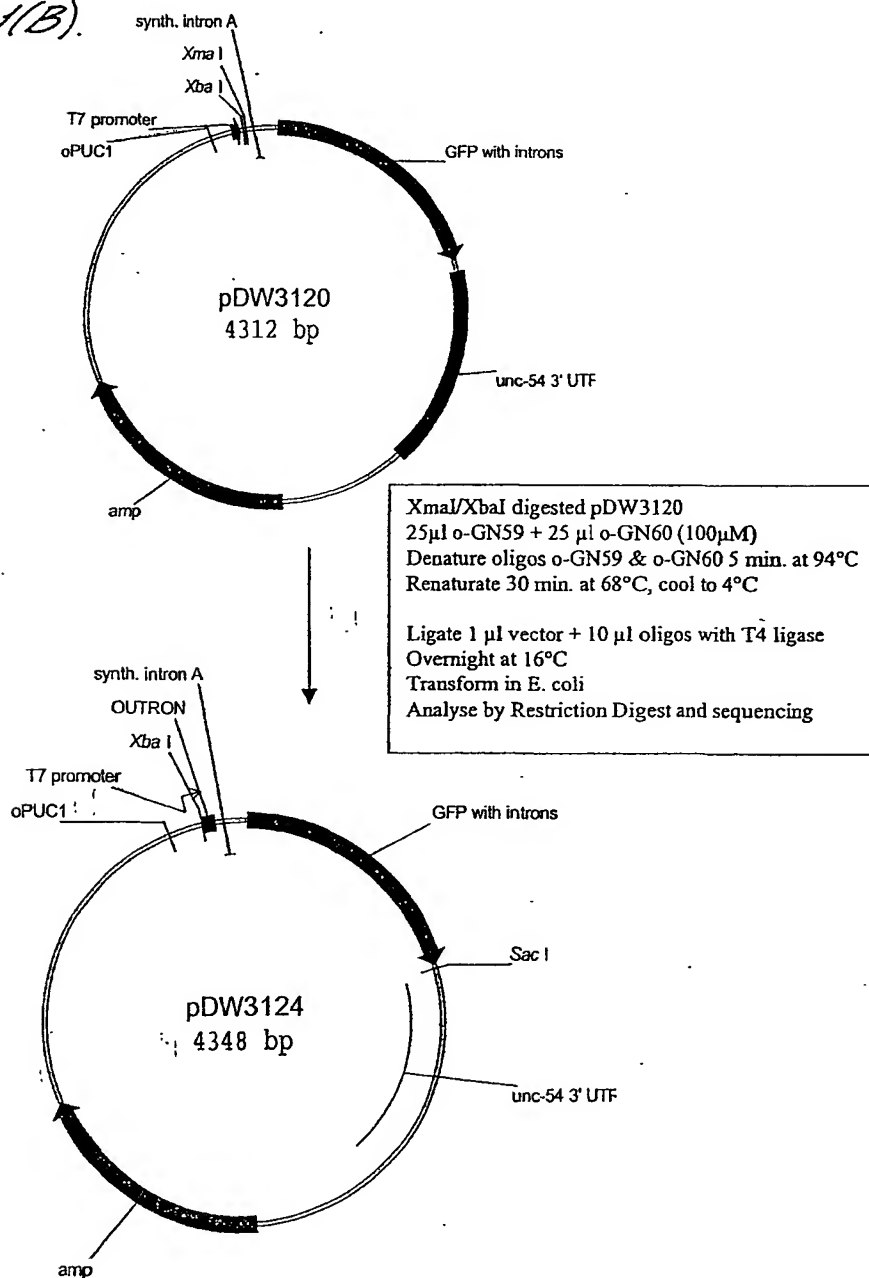




FIG. 2.

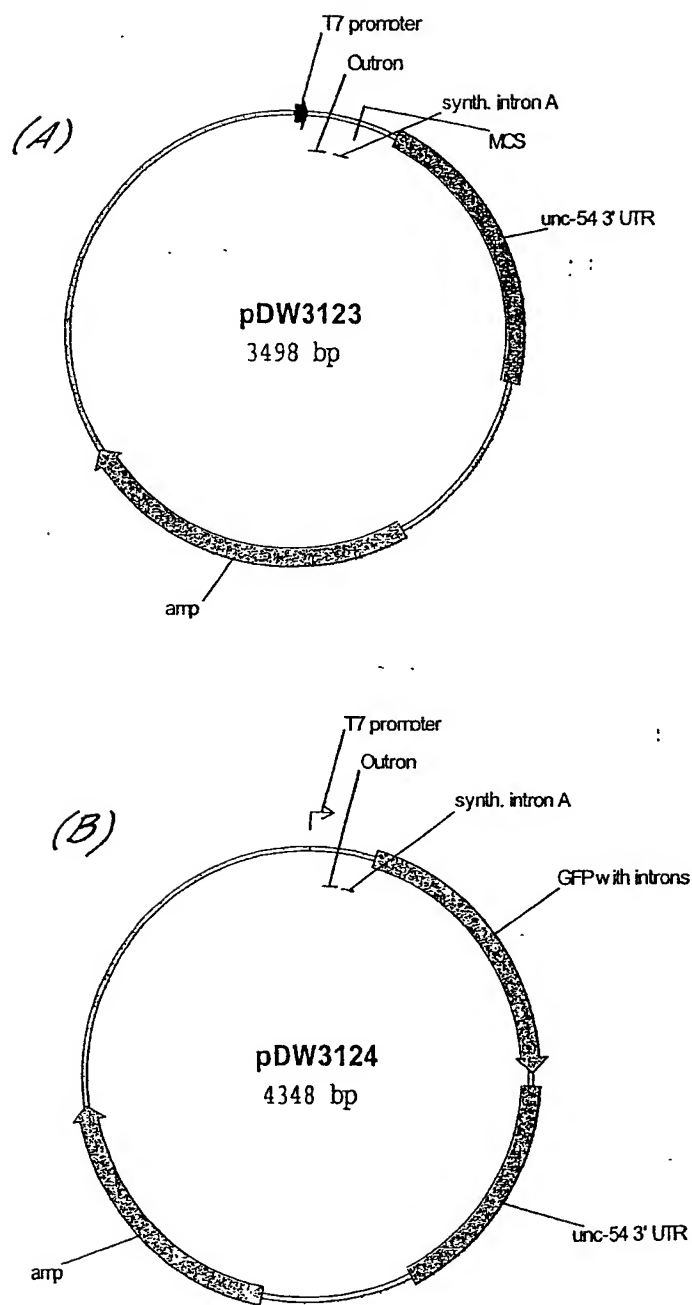
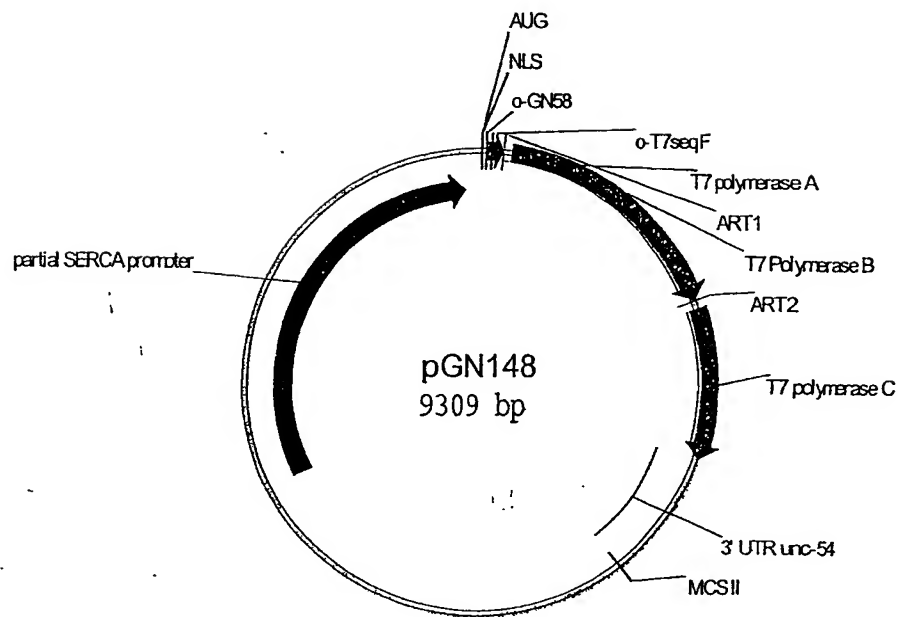


FIG. 3.



*FIG. 4.*

Nucleotide sequence of pGN148

atgactgctccaaagaagaagcgtaagggtaccggtaatgaacacgattaacatcgctaagaacgacttctc  
tgacatcgaaactggctgctatcccggttcaacactctggctgaccattacgggtgagcggttagctcggtaag  
tttaaacatctagatactaactaacgattaacatttaaattttcagcgaacaggtggcccttgagcatgag  
tcttacgagatgggtgaagcacgcttccgcaagatgtttgagcgctcaacttaagctgggtgaggttgcgga  
taacgctgccgccaagcctctcatcactaccctactccctaagatgattgcacgcatcaacgactgggttg  
aggaagtgaagctaaagcgcggaagcgcccgacagccttcagttcctgcaagaaatcaagccggaagcc  
gtagcgtagacatcaccataaagaccactctggcttgccctaaccagtgctgacaatacaaccggttcaggctgt  
agcaagcgcaatcggtcggggccattgaggacgagggctcgcttcggctcgatccgtgaccttgaagctaaagc  
acttcaagaaaaacgttgaggacaactcaacaagcgctagggcacgctctacaagaaagcatttatgcaa  
gttgcgcgaggtgacatgctcttaagggtctactcgggtggcgagggcggtggtcttcgctggcataagggaaga  
ctctattctatgtaggagtagcgtgcatcgagatgctcattgagtcacccggagtggttagcttacaccgccc  
aaaatgctggcgtagtaggtcaagactctgagactatcgaaactcgacacctgaatacgcctgaggttatcgca  
accgctgcaggtgctgctggctggcatctctccgatgttccaaccttgcgtagttcctcctaagccgtggac  
tggcatctactgggtgggtggctattgggttaacggctcgctcctctggcgctgggtgctactcacagtaaga  
aagcactgatgcgctacgaagacgtttacatgcctgaggtgtacaagcgattaaacttgccgcaaaacacc  
gcattggaaaaatcaacaagaaagtcctagcgggtcgccaacgtaatacacaagtgggaagcattgtccggtcga  
ggacatccctgcgattgagcgtgaagaactcccgatgaaacgggaagacatcgacatgaatcctgaggtctc  
tcaccgctggaaacgtgctgcccgtgctgtgtaccgcaaggacaaggctcgcaagctctcgccgtatcagc  
cttgagttcatgcttgagcaagccaataagtttgcataaccataaggccatctgggttcccttacaacatgga  
ctggcgcggtcggtgtttacgctgtgtcaatgttcaaccgcaagctaacgatatgaccaaaggactgctta  
cgctggcgaaaggtaaaccaatcggtaagggaaggttactactggctgaaaaatccacgggtgcaaacgtgctg  
gggtgcgataagggttccgttccctgagcgcacatcaagttcattgaggaaaaccacgagaacatcatggcttg  
cgctaagctctccactggagaacacttgggtgggtgagcaagattctccgttctgcttccctgcttctgct  
ttgagtacgctgggttacagcaccacggcctgagctataactgctcccttccgctggcggtttgacgggtct  
tgctctggcatccagcacttctccgcgatgctccgagatgaggtaggtgggtcgcgcggttgtaagtttaaa  
ctctatcctactaactaacgaagccttatttaaattttcagaacttgcttccctagtgaaacggttcaggaca  
tctacgggattgtttgctaagaagtcacacgagattctacaagcagacgcaatcaatgggaccgataacgaa  
gtagtaccgctgaccgatgagaacactgggtgaaatctctgagaaagtcgaagctgggcactaaaggcactggc  
tggtaaatggctggcttacgggtgttactcgcagtggtgactaagcgttcagtcacgctgggttacgggt  
ccaaagagttcgggttccgttccgtcaacaagtgctggaagataccattcagccagctattgattccggcaagggt  
ctgatgttcaactcagccgaatcaggctgctggatacatggctaaagctgatttgggaatctgtgagcgtgac  
gggtgtagctcggttgaaagcaatgaactggcttaagctgctgctaaagctgctggctgctgaggtcaaaag  
ataagaagactggagagattcttcgcaagcgttgccgtgtgcatgggtcactccggatgggttccctgtg  
tggcaggaatacagaagcctattcaaacgcgtttgaacctgatttccctgggtcagttccgctacagcc  
taccatatacaccacaagaatagcgagattgatgcacacaaacaggagctgggtatcgctcctaactttg  
tacacagccaagacggttagccaccttcgtaagactgtagtgtgggcacacgagaagtcggaatcgaatct  
tttgactgatttcacgactccttcgggtaccattccgggtgacgctgcgaacctgttcaaagcagtgccgga  
aactatgggtgacacatatgagtccttgtagtgactggctgatttctacgaccagttcgctgaccagttgc  
acgagttcctaattggacaaaatgccagcacttccggctaaaggtaacttgaacctccgtgacatcttagag  
tcggacttcgcttcgctgaagaattccaactgagcgccgggtcgctaccattaccaacttgtctgggtgca  
aaaaataataggggcccgtgtcatcagagtaagtttaaactgagttctactaactaacgagtaataatttaa  
ttttcagcatctcgccccgtgcctctgacttctaagtcgaattactcttcaacatccctacatgctcttt  
ctccctgtgctcccacccctatttttggtattatcaaaaaaacttcttctaatttctttgttttttagc  
ttcttttaagtcaccttcaacaatgaaatttgtagattcaaaaatagaatttaattcgtaataaaaaagtcg  
aaaaaaattgtgctccctccccccatttaataataattctatcccaaatctacacaaatgttctgtgtacac  
ttcttatgttttttttacttctgataaaatttttttgaacatcatagaaaaaacgcacacaaaatacct  
tatcatatgttacgtttcagtttatgaccgcaatttttatttcttcgcacgtctgggcccctcatgacgctc  
aatcatgctcatcgtaaaaagttttggagatttttggaaatttttcaatcaagtgaagtttatgaaat  
taattttctgtcttttctgttttttgggggttttccctattgtttgtcagaagttttcagggacggcggttttc  
ttgctaaaaatcacaagttatgatgagcagatgcaagaagatcggaagaaggtttgggtttgaggtcag  
tggaaggtgagtagaagttgataatttgaagtgagtagtgctatgggggtttttgccttaaatgacaga  
atacatcccaatataccaaacataactgtttcctactagtcggccgtacgggcccctttcgctcgcgcgt  
ttcgggtgatgacggtgaaaacctctgacacatgcagctcccggagacgggtcacagcttgtctgtaagcggga  
tgccgggagcagacaagcccgctcagggcgctcagcggtgttgccgggtgtcggggtgtcggttgaactatg  
cggtcatcagagcagattgtactgagagtgacacatctcggtgtgaaataccgcacagatgctgaaggaga

*FIG. 4 (CONTINUED 1.)*

aaataccgcatcaggcgcccttaagggcctcgtgatacgcctatTTTTtataggttaatgtcatgataataa  
tggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaa  
atacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaa  
gagtatgagatttcaacatttccgctgctgcgccttattccctTTTTtgccgcatTTTgccttctgttttg  
ctcaccagaaacgctgggtgaaagttaaagatgctgaagatcagttgggtgcacgagtggttacatcgaa  
ctggatctcaacagcggttaagatccttgagagttttcgcccggaagacggttttccaatgatgagcacttt  
taaagttctgtatgtggcgcggtattatcccgatttgacggcggggaagagcaactcggtcgcccatac  
actattctcagaatgacttgggtgagtagtaccagtcacagaaaagcatcttacggatggcatgacagta  
agagaattatgcagtgctgccataaccatgagtgataaactgcccgaacttacttctgaacacgatcgg  
aggaccgaaggagctaacgcgtttttgacacacatgggggatcatgtaactcgccttgatcgttgggaac  
cggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacggtg  
cgcaactattaaactggcgaactacttactctagcttcccggaacaattaatagactggatggaggcgga  
taaagttgcaggaccacttctgcgctcgcccttccggctggctggtttattgctgataaatctggagcgg  
gtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgatcgtagttatc  
tacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgtgagataggtgcctcactgat  
taagcatttgtaactgtcagaccaagtttactcatatatacttttagattgatttaaaacttcattttta  
ttaaaaggatctagggtgaagatccttttgataatctcatgacaaaatcccttaacgtgagtttctgctc  
cactgagcgtcagaccctgtagaaaagatcaaggatcttcttgagatccttttttctgcgctaatctg  
ctgcttgcaaaaaaaaccaccgctaccagcgggtggtttggttgccggatcaagagctaccaactcctt  
ttccgaaggttaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggc  
caccacttcaagaactctgtagcaccgctacatacctcgtctgtaactcctgttaccagtggtcgtgc  
cagtgcgataaagtcgtgcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgg  
gctgaacggggggttcgtgcacacagcccagcttgagcgaaacgacctacaccgaactgagatacctacag  
cgtgagcatttgaaaagcgccacgcttcccgaaaggagaaaaggcgacaggtatccggttaagcggcaggg  
cggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggttct  
gccacctctgacttgagcgtcgatttttgtagtgcctcgtcagggggcgagcctatggaaaaacgacgc  
aacgcggcctttttacgggttccctggccttttgetggccttttgcctacatgttcttctcgttatcccc  
tgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgcgcagccgaacgacccgagc  
gcagcagtgctcagtgagcgaggaaagcggaagagcgcacaaatcgcaaacgcctctcccgcgcgcttggccg  
attcattaatcagctggcagcagaggttcccgactggaaagcgggcagtgagcgcgaacgcaattaatgt  
gagttagctcactcattaggcaccacagcctttacactttatgcttccggctcgatagtgtgtggaattg  
tgagcggataacaatttcacacaggaacacagctatgacatgattacgccaagctgtaagttaaacatga  
tcttactaactaactattctcatttaaattttcagagcttaaaaatggctgaaatcactcacaacgatgga  
tacgctaacaacttggaatgaaataagcttgcatgcctgcagagcaaaaaatactgcttttcttgcaa  
aattcgggtgctttcttcaaagagaaacttttgagtcggcgcgagcatttcttcttgacttctctctt  
ccgcaaaaaagcctagcatttttattgataatttgattacacacactcagagttcttcgacatgataagt  
gtttcattggcactcgcctaacagtagacagaaagggcggtatttatcgatcgatattgaagacaaact  
ccaaatgtgtctcattttggagccccgtgtggggcagctgctctcaatatattactaggagagcaggag  
ggggaccttatcgaaacgtcgcagcagcatttcttcttcttatgactctcttctcactctctcacacatta  
atcgattcatagactccarattccttgatgaagtggtggttttttagcttttttcccgatttgtaaaag  
gaagagggtgacgatgttaggaaaaagagaacggagccgaaaaacatccgtagtaagtcttcttttaag  
ccgacacttttttagacagcattcgcgcgtagttttgaagtttaaattttaaaaaataaaaattagtttcaa  
tttttttaattactaaataggcaaaagttttttcaagaactctagaaaaactagcttaattcatgggtac  
tagaaaaattctgttttaatttaatatcttaagatgtaattacgagaagcgtttttgaaaaattct  
caattaaaagaatttgccgatttagaataaaagcttccagaaatgagtaaaagctcaaatagaagttgt  
ttttaaggaaaaaacgaaaaaagaacactattatcttttccctcccgctaaaaattagttgtgtgat  
aatagtgatccgctgtctatttgactcggctcttcacaccgtgcttctctcacttgacccaacaggaaa  
aaaaaacatcacgtctgagacggtgaattgccttatcaagacgctcgtctcttccaccagtaacaaaaaa  
aatttggtttcttactttatatttgtaggtcacaacaaaaaagtgatgagttttgtgggtcggttgg  
ctccacaccacctccgctccagcagcacacaatcatcttctgtgtgtctcgcagattccttgatgcgc  
ggctggaatgcacacattcgcgcgcgaactacacaccacactcactttcgggtggtattactacacgtca  
tcgtgtgtcgtagtctccgctcttctgctcccaactcactcctcattatcccttgggtgattgatttt  
tttaaatggtacaccactcctgacgtttctacettctgttttccgtccatttagatttatctggaaatt  
tttttaaaatttttagccagagagttctagtcttctgttctaaaagtctaggtcagacatacattttctatt  
tctcatcaaaaaaagttagataaaagaaactggttattcagaaagagtgctcctcgttgaaattgattca  
aaaaaaattccccaccctcgttggtttctcaaaatagagatcaacgggaatttttcttctcgtattcaat  
tttttgctgcgctctgtctgcaaaagtgtgtgtccgagcaaaagatgagagaatttacaacagaaatg

FIG. 4 (CONTINUED 2.)

aaaaaaagtgtggccaaataatgaagttttatccgagattgatgggaaagatattaatgttctttacgggtt  
 ggaggggagagagagatagattttcgcatacaactccgccttttacatgtcttttagaatctaaaatagat  
 ttttctcatcatttttaatatagaaaatcgagaataattacagtaatttttcgaattttcttgccaaaaatacacg  
 aaattgttgggtctcgccagatctcggtcttagtgggtcatttgggtttaaagtttataaaatttccaaat  
 tctagtgttttaattttccgcataaattggacctaaaatgggtttttgtgcatcattttcaacaagaatacgtga  
 aaatcctgttgttttcgaattttcttttcaaaaaatacacgaaatatatggtaatttcccgaaatatttgagg  
 gtctcgccacgatttccagtccagtgggcaggatttatcacgaaaaaagttcgcttagtctcacatttcg  
 gaaaaaccgaattcaaattagttttttgtcatcattttgaacaaaaaatcgagacatccctatagtttcgca  
 attttcgtcgctttttctccaaaaatgcagctcagaattaaaattcgtcggaactgggaccatgatatac  
 tttctccccgtttttcatttttattttttttatctacatggattgactaaaggctaccaccaccgcgagtgtg  
 tggcatatcacacacacacacacacacacaaatgtcgagattttatgtgttatccctgcttgatttcgttcggt  
 tgtctctctctctctattcatcttttgagccgagaagctccagagaatggagcacacaggatcccgccgcg  
 cgatgtcgtcgggagatggcgccgcctgggaagccgcgagagatatcagggaagatcgtctgatttctcc  
 tggatgccacctcatctcagttttctccgctgttactccctgcgaactgatatcttccgcttgtcgt  
 taagagatgttttttatttttttaccacgggtctctctctctcgccacagacgtcagttgttggctgtg  
 tgtcgggtctctgccacggcgggcctcatcttcttcttcttctctctctctctctctctctctctctctct  
 ctattcattcttattccttttcatcatcaaaactagcatttcttactttattttttttcaattttcaat  
 tttcagataaaaaccaaactacttgggttacagccgtcaacagatccccgggattggccaaaggacccaaag  
 gtagtgttccgaatgatactaacataacatagaacattttcaggaggacccttgcttggaggggtaccgagct  
 cagaaaaa

*FIG. 5.*

```

=====
T7 promoter                                     Outron
=====
1  AGCTTGGCGC CTAATACGAC TCACTATAGG GCTGCAGGTC GACTCTAGAT TACAACATAAT TATACTTATT
   TCGAACC CGG GATTATGCTG AGTGATATCC CGACGTCCAG CTGAGATCTA ATGTTGATTA ATATGAATAA

Outron
=====
71  TGAATATTCA AATTTTCAGA CCCGGGATTG GCCAAAGGAC CCAAAGGTAT GTTTCGAATG ATACTAACAT
    ACTTATAAGT TTAAGTCT GGGCCCTAAC CGGTTTCCTG GGTTCCTATA CAAAGCTTAC TATGATTGTA

synth. intron A                               MCS
=====
141 AACATAGAAC ATTTTCAGGA GGACCCTTGG CTAGCGTCCT GCTGGGATTA CACATGGCAT GGATGAACTA
    TTGTATCTTG TAAAGTCTT CCTGGGAACC GATCGCAGGA CGACCCTAAT GTGTACCGTA CCTACTTGAT

=====
unc-54 3' UTR
=====
211 TACAAATAGG GCCGGCCGAG CTCGCGATCG GCCGCTGTCA TCAGATCGCC ATCTCGCGCC CGTGCCTCTG
    ATGTTTATCC CGGCCGGCTC GAGGCGTAGC CGGCGACAGT AGTCTAGCGG TAGAGCGCGG GCACGGAGAC

=====
unc-54 3' UTR
=====
281 ACTTCTAAGT CCAATTACTC TTCAACATCC CTACATGCTC TTTCTCCCTG TGCTCCACCC CCCTATTTTT
    TGAAGATTCA GGTTAATGAG AAGTTGTAGG GATGTACGAG AAAGAGGGAC ACGAGGGTGG GGGATAAAAA

=====
unc-54 3' UTR
=====
351 GTTATTATCA AAAAACTTC TTCTTAATTT CTTGTGTTTT TAGCTTCTTT TAAGTCACCT CTAACAATGA
    CAATAATAGT TTTTGTGAAG AAGAATTAAA GAAACAAAAA ATCGAAGAAA ATTCAAGTGA GATTGTACT

=====
unc-54 3' UTR
=====
421 AATTGTGTAG ATTCAAAAAA AGAATTAATT CGTAATAAAA AGTCGAAAAA AATTGTGCTC CCTCCCCCA
    TTAACACATC TAAGTTTTTA TCTTAATTAA GCATTATTTT TCAGCTTTTT TTAACACGAG GGAGGGGGGT

=====
unc-54 3' UTR
=====
491 TTAATAATAA TTCTATCCCA AAATCTACAC AATGTTCTGT GTACACTTCT TATGTTTTTT TTAATCTCTGA
    AATTATTATT AAGATAGGGT TTAGATGTG TTACAAGACA CATGTGAAGA ATACAAAAAA AATGAAGACT

=====
unc-54 3' UTR
=====
561 TAAATTTTTT TTGAAACATC ATAGAAAAAA CCGCACACAA AATACCTTAT CATATGTTAC GTTTCAGTTT
    ATTTAAAAAA AACTTTGTAG TATCTTTTTT GCGGTGTGTT TTATGGAATA GTATACAATG CAAAGTCAAA

unc-54 3' UTR
=====
631 ATGACCGCAA TTTTATTTT TTCGCACGTC TGGGCTCTC ATGACGTCAA ATCATGCTCA TCGTGAAAAA
    TACTGGCGTT AAAAAATAAG AAGCGTGCAG ACCCGGAGAG TACTGCAGTT TAGTACGAGT AGCACTTTTT

unc-54 3' UTR
=====
701 GTTTTGGAGT ATTTTGGAA TTTTCAATC AAGTGAAAGT TTATGAAATT AATTTTCCTG CTTTGTCTTT
    CAAAACCTCA TAAAAACCTT AAAAAGTTAG TTCACCTTCA AATACTTTAA TTAAPAGGAC GAAAACGAA

unc-54 3' UTR
=====
771 TTGGGGGTTT CCCCTATTGT TTGTCAAGAG TTTCGAGGAC GGCGTTTTTC TTGCTAAAAT CACAAGTATT
    AACCCCAAAA GGGGATAACA AACAGTTCTC AAAGCTCCTG CCGCAAAAAG AACGATTTTA GTGTTCTATA

unc-54 3' UTR
=====
841 GATGAGCAGC ATGCAAGAAA GATCGGAAGA AGGTTTGGGT TTGAGGCTCA GTGGAAGGTG AGTAGAAGTT
    CTACTCGTGC TACGTTCTTT CTAGCCTTCT TCCAAACCCA AACTCCGAGT CACCTCCAC TCATCTTCAA

unc-54 3' UTR
=====
911 GATAATTGTA AAGTGGAGTA GTGTCTATGG GGTTTTGCC TTAATGACA GAATACATTC CCAATATACC
    CTATTAACT TTCACCTCAT CACAGATACC CAAAAACGG AATTACTGT CTTATGTAAG GGTATATGG

unc-54 3' UTR
=====
981 AAACATAACT GTTTCCTACT AGTCGGCGGT ACGGGCCCTT TCGTCTCGCG CGTTTCGGTG ATGACGGTGA
    TTTGTATTGA CAAAGGATGA TCAGCCGGCA TGCCCGGGAA AGCAGAGCGC GCAAAGCCAC TACTGCCACT

```

*FIG. 5 (CONTINUED 1.)*

1051 AAACCTCTGA CACATGCAGC TCCCGGAGAC GGTACAGCT TGTCTGTAAG CGGATGCCGG GAGCAGACAA  
TTTGAGACT GTGTACGTCG AGGGCCTCTG CCAGTGTCTGA ACAGACATTC GCCTACGGCC CTCGTCTGTT

1121 GCGGCTCAGG GCGCGTCAGC GGGTGTGGC GGGTGTGGG GCTGGCTTAA CTATGCGGCA TCAGAGCAGA  
CGGGCAGTCC CGCGCAGTCG CCCACAACCG CCCACAGCCC CGACCGAATT GATACGCCGT AGTCTCGTCT

1191 TTGTACTGAG AGTGCACCAT ATGCGGTGTG AAATACCGCA CAGATGCGTA AGGAGAAAAT ACCGCATCAG  
AACATGACTC TCACGTGGTA TACGCCACAC TTTATGGCGT GTCTACGCAT TCCTCTTTTA TGGCGTAGTC

1261 GCGGCCTTAA GGGCCTCGTG ATACGCCTAT TTTTATAGGT TAATGTCATG ATAATAATGG TTTCTTAGAC  
CGCCGGAATT CCCGGAGCAC TATGCGGATA AAAATATCCA ATTACAGTAC TATTATTACC AAAGAATCTG

1331 GTCAGGTGGC ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTTCTAAAT ACATTCAAAT  
CAGTCCACCG TGAAAAGCCC CTTTACACGC GCCTTGGGGA TAAACAAATA AAAAGATTTA TGTAAGTTTA

amp  
=====

1401 ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG AGTATGAGTA  
TACATAGGCG AGTACTCTGT TATTGGGACT ATTTACGAAG TTATTATAAC TTTTCTCTTC TCATACTCAT

amp  
=====

1471 TTCAACATT CCGTGTGCGC CTTATTCCT TTTTTCGGC ATTTTGCCTT CCTGTTTTG CTCACCCAGA  
AAGTTGTAAA GGCACAGCGG GAATAAGGGA AAAAACGCGG TAAACGGAA GGACAAAAAC GAGTGGGTCT

amp  
=====

1541 AACGCTGGTG AAAGTAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC  
TTGCGACCAC TTTCATTTT TACGACTTCT AGTCAACCCA CGTGCTCACC CAATGTAGCT TGACCTAGAG

amp  
=====

1611 AACAGCGGTA AGATCCTTGA GAGTTTTCGC CCCGAAGAAC GTTTTCCAAT GATGAGCACT TTAAAGTTT  
TTGTGCCCAT TCTAGGAACT CTCAAAAGCG GGGCTTCTTG CAAAAGGTTA CTACTCGTGA AAATTTCAAG

amp  
=====

1681 TGCTATGTGG CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTGCGCGCA TACACTATT  
ACGATACACC GCGCCATAAT AGGCATAAC TGCGGCCCGT TCTCGTTGAG CCAGCGGCGT ATGTGATAAG

amp  
=====

1751 TCAGAATGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTTACGG ATGGCATGAC AGTAAGAGAA  
AGTCTTACTG AACCAACTCA TGAGTGGTCA GTGTCTTTT GTAGAATGCC TACCGTACTG TCATTCTCTT

amp  
=====

1821 TTATGCAGTG CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG ATCGGAGGAC  
AATACGTCAC GACGGTATTG GTACTACTA TGTGACGCC GGTGGAATGA AGACTGTTGC TAGCCTCCTG

amp  
=====

1891 CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACGCG CTTGATCGTT GGAACCGGA  
GCTTCTCTGA TTGGCGAAAA AACGTGTTGT ACCCCCTAGT ACRTTGAGCG GAACTAGCAA CCCTTGGCCT

amp  
=====

1961 GCTGAATGAA GCCATACCAA ACGACGAGCG TGACACCACG ATGCCTGTAG CAATGGCAAC AACGTTGCCG  
CGACTTACTT CGGTATGGTT TGCTGCTCGC ACTGTGGTGC TACGGACATC GTTACCGTTG TTGCAACGCG

*FIG. 5 (CONTINUED 2.)*

amp  
=====

2031 AAACATTAAT CTGGCGAAGT ACTTACTCTA GCTTCCCGGC AACAATTAAT AGACTGGATG GAGGCGGATA  
TTTGATAATT GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC CTCCGCCTAT

amp  
=====

2101 AAGTTGCAGG ACCACTTCTG CGCTCGGCC TCCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG  
TTCAACGTCC TGGTGAAGAC GCGAGCCGGG AAGGCCGACC GACCAAATAA CGACTATTTA GACCTCGGCC

amp  
=====

2171 TGAGCGTGGG TCTCGCGSTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC  
ATTCGCACCC AGAGCGCCAT AGTAACGTCG TGACCCCGGT CTACCATTCC GGAGGGCATA GCATCAATAG

amp  
=====

2241 TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT GCCTCACTGA  
ATGTGCTGCC CTCAGTCCG TTGATACCTA CTGCTTTTAT CTGCTAGCG ACTCTATCCA CGGAGTGACT

amp  
=====

2311 TTAAGCATTT GTAACGTGCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAC TTCATTTTAA  
AATTCGTAAC CATTGACAGT CTGGTTCAAA TGAGTATATA TGAAATCTAA CTAAATTTTG AAGTAAAAAT

2381 ATTTAAAGG ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTCG  
TAAATTTTCC TAGATCCACT TCTAGGAAA ACTATTAGAG TACTGGTTTT AGGGAATTGC ACTCAAAAGC

2451 TTCCACTGAG CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA  
AAGGTGACTC GCAGTCTGGG GCATCTTTT TAGTTTCTTA GAAGAACTCT AGGAAAAAAA GACGCGCATT

2521 TCTGCTGCTT GCAACAAAAA AAACCACCGC TACCAGCGGT GGTGTTGTTG CCGGATCAAG AGCTACCAAC  
AGACGACGAA CGTTGTTTT TTTGGTGGCG ATGGTCGCGA CCAAACAAAC GGCCTAGTTC TCAGATGGTTG

2591 TCTTTTTCCG AAGGTAACGT GCTTCAGAGC AGCGCAGATA CCAATACTG TCCTTCTAGT GTAGCCGTAG  
AGAAAAAGGC TTCCATTGAC CGAAGTCGTC TCCTGCTTAT GGTTTATGAC AGGAAGATCA CATCGGCATC

2661 TTAGGCCACC ACTTCAAGAA CTCTGTAGCA CCGCTACAT ACCTCGCTCT GCTAATCCTG TTACCACTGG  
AATCCGGTGG TGAAGTTCTT GAGACATCGT GCGCGATGTA TGGAGCGAGA CGATTAGGAC AATGGTCACC

2731 CTGCTGCCAG TGGCGATAAG TCGTGCTTTA CCGGGTTGGA CTCAAGACGA TAGTTACCGG ATAAGGCGCA  
GACGACGGTC ACCGCTATTC AGCACAGAAT GGGCCAACT GAGTTCTGCT ATCAATGGCC TATTCCGCGT

2801 GCGGTGCGGC TGAACGGGGG GTTCGTGCAC ACAGCCCAGC TTGGAGCGAA CGACCTACAC CGAACTGAGA  
CGCCAGCCCC ACTTGCCCC CAAGCACGTG TGTCGGGTCG AACCTCGCTT GCTGGATGTG GCTTGACTCT

2871 TACCTACAGC GTGAGCATTT AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GCGGACAGG TATCCGGTAA  
ATGGATGTCG CACTCGTAAC TCTTTCGCGG TGCGAAGGGC TTCCCTCTTT CCGCTGTGCC ATAGGCCATT

2941 GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGGAAC GCCTGGTATC TTTATAGTCC  
CGCCGTCCCA GCCTGTCTCT CTCGCGTGCT CCCTCGAAGG TCCCCCTTG CCGACCATAG AAATATCAGG

3011 TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTGG TGATGCTCGT CAGGGGGGCG GAGCCTATGG  
ACAGCCCAA GCGGTGGAGA CTGAATCGC AGCTAAAAAC ACTACGAGCA GTCCCCCGC CTGCGATACC

3081 AAAAACGCCA GCAACGCGGC CTTTTTACGG TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC ATGTTCTTTC  
TTTTTGCGGT CGTTGCGCG GAAAAATGCC AAGGACCGGA AAACGACCGG AAAACGAGTG TACAAGAAAG

3151 CTGCGTTATC CCCTGATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACCG CTCGCCGAG  
GACGCAATAG GGGACTAAGA CACCTATTGG CATAATGGCG GAAACTCACT CGACTATGGC GAGCGCGCTC

3221 CCGAACGACC GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCGCC CAATACGCAA ACCGCCTCTC  
GGCTTGCTGG CTCGCGTCGC TCAGTCACTC GCTCCTTCGC CTTCTCGCGG GTTATGCGTT TGGCGGAGAG

3291 CCCGCGGCTT GGCCGATTCA TTAATGCAGC TGGCACGACA GGTTCCTCGA CTGGAAAGCG GGCAGTGAGC  
GGGCGCGCAA CCGCTAAGT AATTACGTCG ACCGTGCTGT CCAGAGGGCT GACCTTTCGC CCGTCACTCG

3361 GCAACGCAAT TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTTA CACTTTATGC TTCGGGCTCG  
CGTTGCGTTA ATTACACTCA ATCGAGTGAG TAATCCGTGG GGTCCGAAAT GTGAAATACG AAGGCCGAGC

3431 TATGTTGTGT GGAATTGTGA GCGGATAACA ATTCACACA GGAAACAGCT ATGACCATGA TTACGCCA  
ATACAACACA CCTTAACACT CGCCTATTGT TAAAGTGTGT CTTTGTCTGA TACTGGTACT AATGCGGT



*FIG. 6.*

```

T7 promoter                                     Outron
=====
1  AGCTTGGCGC CTAATACGAC TCACTATAGG GCTGCAGGTC GACTCTAGAT TACAACAAAT TATACTTATT
   TCGAACCGCG GATTATGCTG AGTGATATCC CGACGTCCAG CTGAGATCTA ATGTTGATTA ATATGAATAA

Outron                                     synth. intron A
=====
71  TGAATATTCA AATTTTCAGA CCCGGGATTG GCCAAAGGAC CCAAAGGTAT GTTTCGAATG ATAGTAACAT
    ACTTATAAGT TTAAAAGTCT GGGCCCTAAC CGGTTTCCTG GGTTCCTATA CAAAGCTTAC TATGATTGTA

synth. intron A                                     GFP with introns
=====
141 AACATAGAAC AATTTTCAGG GACCCCTTGG CTAGCGTCGA CGGTACCATG GGGCGCGCCA TGAGTAAAGG
    TTGTATCTTG TAAAAGTCTT CCTGGGAACC GATCGCAGCT GCCATGGTAC CCCGCGCGGT ACTCATTTC

GFP with introns
=====
211 AGAAGAACTT TCACTGAGG TTGTCCCAAT TCTTGTTGAA TTAGATGGTG ATGTTAATGG GCACAAATTT
    TCTTCTTGAA AAGTGACCTC AACAGGGTTA AGAACAACTT AATCTACCAC TACAATTACC CGTGTTTAA

GFP with introns
=====
281 TCTGTACAGT GAGAGGGTGA AGGTGATGCA ACATACGGAA AACTTACCCT TAAATTTATT TGCCTACTG
    AGACAGTCAC CTCTCCCACT TCCACTACGT TGTATGCCTT TTGAATGGGA ATTAAATAA ACGTGATGAC

GFP with introns
=====
351 GAAACTACCC TGTTCCATGG GTAAGTTTAA ACATATATAT ACTAACTAAC CCTGATTATT TAAATTTTCA
    CTTTGATGAG ACAAGGTACC CATTCAAATT TGTATATATA TGATTGATTG GGAATAATAA ATTTAAAGT

GFP with introns
=====
421 GCAACAACTT GTCACACTT TCTGTTATGG TGTTCAATGC TTCTCGAGAT ACCCAGATCA TATGAAACGG
    CGGTTGTGAA CAGTGATGAA AGACAATACC ACAAGTTACG AAGAGCTCTA TGGGTCTAGT ATACTTTGCC

GFP with introns
=====
491 CATGACITTT TCAAGAGTGC CATGCCCGAA GGTATGTAC AGGAAAGAAC TATATTTTC AAAGATGACG
    GTACTGAAA AGTTCTCAGC GTACGGGCTT CCAATACATG TCCTTTCTTG ATATAAAAG TTTCTACTGC

GFP with introns
=====
561 GGAACACAA GACACCTAAG TTTAAACAGT TCGGTACTAA CTAACCATAC ATATTTAAAT TTTGAGGTGC
    CCTTGATGTT CTGTGCATTC AAATTTGTCA AGCCATGATT GATTGGTATG TATAAATTA AAAGTCCAGC

GFP with introns
=====
631 TGAAGTCAAG TTTGARGGTG ATACCCITGT TAATAGAATC GAGTTAAAG GTATTGATTT TAAAGAAGAT
    ACTTCAGTTC AACTTCCAC TATGGGAACA ATTATCTTAG CTCAATTTTC CATAACTAAA ATTTCTTCTA

GFP with introns
=====
701 GGAACATTC TTGGACACAA ATTGGAATAC AACTATAACT CACACAATGT ATACATCATG GCAGACAAAC
    CCTTGTGAAG AACCTGTGTT TAACCTTATG TTGATATTGA GTGTGTACA TATGTAGTAC CGTCTGTTG

GFP with introns
=====
771 AAAAGAATGG AATCAPAGTT GTAAGTTTAA ACTTGGACTT ACTAACTAAC GGATTATATT TAAATTTTCA
    TTTCTTACC TTAGTTTCAA CATTCAAATT TGAACCTGAA TGATTGATTG CCTAATATAA ATTTAAAGT

GFP with introns
=====
841 GAACTTCAAA ATTAGACACA ACATTGAAGA TGAAGCGTT CAACTAGCAG ACCATTATCA ACAAAATACT
    CTGAAGTTT TAATCTGTGT TGAACCTCT ACCTTCGCAA GTTGATCGTC TGGTAATAGT TGTTTTATGA

GFP with introns
=====
911 CCAATTGGCG ATGGCCCTGT CCTTTTACCA GACAACCATT ACCTGTCCAC ACAATCTGCC CTTTCGAAAG
    GGTTAACCGC TACCGGACAA GGAATATGGT CTGTTGTAA TGGACAGGTG TGTTAGACGG GAAAGCTTTC

GFP with introns
=====
981 ATCCCAACGA AAAGAGAGAC CACATGGTCC TTCTTGAGTT TGTAACAGCT GCTGGGATTA CACATGGCAT
    TAGGGTTGCT TTTCTCTCTG GTGTACCAGG AAGAATCAA ACATTGTGCA CGACCTAAT GTGTACCGTA

```

*FIG. 6 (CONTINUED 1.)*

GFP with introns

unc-54 3' UTR

1051 GGATGAACTA TACAAATAGG GCCGGCCGAG CTCGCGATCG GCCGCTGTCA TCAGATCGCC ATCTCGCGCC  
CCTACTTGAT ATGTTTATCC CGGCCGGCTC GAGCGGTAGC CGGCGACAGT AGTCTAGCGG TAGAGCGCGG

unc-54 3' UTR

1121 CGTGCCCTCTG ACTTCTAAGT CCAATTACTC TTCAACATCC CTACATGCTC TTTCTCCCTG TGCTCCACCC  
GCACGGAGAC TGAAGATTCA GGTTAATGAG AAGTTGTAGG GATGTACGAG AAAGAGGGAC ACAGGGGTGG

unc-54 3' UTR

1191 CCTATTTTT GTTATTATCA AAAAACTTC TTCTTAATTT CTTGTGTTTT TAGCTTCTTT TAAGTCACCT  
GGGATAAAAA CAATAATAGT TTTTGTGAAG AAGAATTAAA GAAACAAAA ATCGAAGAAA ATTCAAGTGA

unc-54 3' UTR

1261 CTAACTATGA AATTGTGTAG ATTCAAAAT AGAATTAATT CGTAATAAAA AGTCGAAAAA AATTGTGCTC  
GATTGTTACT TTAACACATC TAAGTTTTTA TCTTAATTAA GCATTATTTT TCAGCTTTTT TTAACACGAG

unc-54 3' UTR

1331 CCTCCCCCA TTAATAATAA TTCTATCCCA AAATCTACAC AATGTTCTGT GTACACTTCT TATGTTTTT  
GGAGGGGGGT AATTATTATT AAGATAGGGT TTTAGATGTG TTACAAGACA CATGTGAAGA ATACAAAAA

unc-54 3' UTR

1401 TTACTTCTGA TAAATTTTT TTGAAACATC ATAGAAAAA CCGCACACAA AATACCTTAT CATATGTTAC  
AATGAAGACT ATTTAAAAA AACTTTGTAG TATCTTTTTT GCGGTGTGTT TTATGGAATA GTATACAATG

unc-54 3' UTR

1471 GTTTCAGTTT ATGACCGCAA TTTTATTTT TTGCGACGTC TGGGCCTCTC ATGACGTCAA ATCATGCTCA  
CAAAGTCAAA TACTGGCGTT AAAAATAAAG AAGCGTGCAG ACCCGGAGAG TACTGCAGTT TAGTACGAGT

unc-54 3' UTR

1541 TCGTGAAGAA GTTTGGAGT ATTTTGGAA TTTTCAATC AAGTGAAAGT TTATGAATTT AATTTTCTG  
AGCACTTTTT CAAAACCTCA TAAAAACCTT AAAAAGTTAG TTAATTTTCA AATACTTTAA TTAAAGGAGC

unc-54 3' UTR

1611 CTTTTGCTTT TTGGGGGTTT CCCCTATTGT TTGTCAAGAG TTTGAGGAC GCGTTTTTC TTGCTAAAAT  
GAAACGAAA AACCCCAAA GGGGATAACA AACAGTTCTC AAAGCTCTG CCGCAAAAG AACGATTTTA

unc-54 3' UTR

1681 CCAAGTATT GATGAGCAG ATGCAAGAAA GATCGGAAGA AGGTTTGGGT TTGAGGCTCA GTGGAAGGTG  
GTGTTTATAA CTACTCGTGC TACGTTCTTT CTAGCCTTCT TCCAAACCCA AACTCCGAGT CACCTCCAC

unc-54 3' UTR

1751 AGTAGAAGTT GATAATTGA AAGTGGAGTA GTGTCTATGG GGTTTTGGC TTAAATGACA GAATACATTC  
TCATCTTCAA CTATTAACT TTCACCTCAT CACAGATACC CCAAAACCG AATTACTGT CTTATGTAAG

unc-54 3' UTR

1821 CCAATATACC AAACATAACT GTTTCCTACT AGTCGGCCGT ACGGGCCCTT TCGTCTCGCG CGTTTCGGTG  
GGTTATATGG TTTGTATTGA CAAAGGATGA TCAGCCGGCA TGCCCGGGAA AGCAGAGCGC GCAAGCCAC

1891 ATGACGGTGA AAACCTCTGA CACATGCAGC TCCCGGAGAC GGTACAGCT TGTCTGTAAG CGGATGCCGG  
TACTGCCACT TTTGGAGACT GTGTACGTCG AGGCGCTCTG CCAGTGTCTG ACAGACATTC GCCTACGGCC

1961 GAGCAGACAA GCCCGTCAGG GCGCGTCAGC GGGTGTGGC GGGTGTGGG GCTGGCTTAA CTATGCGGCA  
CTGCTCTGTT CGGCGAGTCC CGGCGAGTCG CCCACAACCG CCCACAGCCC CGACCGAATT GATACGCCGT

2031 TCAGAGCAGA TTGTACTGAG AGTGACCAT ATGCGGTGTG AAATACCGCA CAGATGCGTA AGGAGAAAAT  
AGTCTCGTCT AACATGACTC TCACGTGGTA TACGCCACAC TTTATGGCGT GTCTACGCAT TCCTCTTTTA

2101 ACGCATCAG GCGGCCTTAA GGGCCTCGTG ATACGCCAT TTTTATAGGT TAATGTCATG ATAATAATGG  
TGCGGTAGTC CGCCGGAATT CCCGGAGCAC TATGCGGATA AAAATATCCA ATTACAGTAC TATTATTACC

## FIG. 6 (CONTINUED 2.)

2171 TTTCTTAGAC GTCAGGTGGC ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGTTTAT TTTCTAAAT  
AAAGAATCTG CAGTCCACCG TGAAAAGCCC CTTTACACGC GCCTTGGGGA TAAACAAATA AAAAGATTTA

2241 ACATTCAAAT ATGTATCCGC TCATGAGACA ATAACCCTGA TAAATGCTTC AATAATATTG AAAAAGGAAG  
TGTAAGTTTA TACATAGGCG AGTACTCTGT TATTGGGACT ATTTACGAAG TTATTATAAC TTTTCTCTTC

=====  
amp  
=====

2311 AGTATGAGTA TTCAACATTT CCGTGTGCC CTTATTCCCT TTTTGGCGC ATTTTGCCTT CCTGTTTTTG  
TCATACTCAT AAGTTGTAA GGCACAGCGG GAATAAGGGA AAAACGCCG TAAACGGAA GGACAAAAC

=====  
amp  
=====

2381 CTCACCCAGA AACGCTGGTG AAAGTAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA  
GAGTGGGTCT TTGCGACCAC TTTCATTTTC TACGACTTCT AGTCAACCCA CGTGCTCACC CAATGTAGCT

=====  
amp  
=====

2451 ACTGGATCTC AACAGCGGTA AGATCCTTGA GAGTTTTCGC CCCGAAGAAC GTTTTCCAAT GATGAGCACT  
TGACCTAGAG TTGTCGCCAT TCTAGGAAGT CTCAAAAGCG GGGCTTCTTG CAAAAGGTTA CTACTCGTGA

=====  
amp  
=====

2521 TTAAAGTTC TGCTATGTGG CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTCGCGGCA  
AAATTTCAAG ACGATACACC GCGCCATAAT AGGGCATAAC TGCGGCCCGT TCTCGTTGAG CCAGCGGCGT

=====  
amp  
=====

2591 TACACTATTC TCAGAATGAC TTGGTTGAGT ACTCACCAGT CACAGAAAAG CATCTTACGG ATGGCATGAC  
ATGTGATAAG AGTCTTACTG AACCAACTCA TGAGTGGTCA GTGTCTTTTC GTAGAATGCC TACCGTACTG

=====  
amp  
=====

2661 AGTAAGAGAA TTATGCAGTG CTGCCATAAC CATGAGTGAT AACACTGCGG CCAACTTACT TCTGACAACG  
TCATTCTCTT AATACGTCAC GACGGTATTG GTACTCACTA TTGTGACGCC GGTGAATGA AGACTGTTGC

=====  
amp  
=====

2731 ATCGGAGGAC CGAAGGAGCT AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAAGTCGC CTTGATCGTT  
TAGCCTCCTG GCTTCTCTGA TTGGCGAAAA AACGTGTTGT ACCCCCTAGT ACATTGAGCG GAACTAGCAA

=====  
amp  
=====

2801 GGGAACCGGA GCTGAATGAA GCCATACCAA ACGACGAGCG TGACACCACG ATGCCTGTAG CAATGGCAAC  
CCCTTGGCCT CGACTTACTT CGGTATGGTT TGCTGCTCGC ACTGTGGTGC TACGGACATC GTTACCGTTG

=====  
amp  
=====

2871 AACGTTGCGC AAACATATTA CTGGCGAACT ACTTACTCTA GCTTCCCGGC AACAAATTAAT AGACTGGATG  
TTGCAACGCG TTTGATAATT GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC

=====  
amp  
=====

2941 GAGGCGGATA AAGTTGCAGG ACCACTTCTG CGCTCGGCC TTCCGGCTGG CTGGTTTATT GCTGATAAAT  
CTCCGCTTAT TTCAACGTCC TGGTGAAGAC GCGAGCCGGG AAGGCCGACC GACCAAATAA CGACTATTTA

=====  
amp  
=====

3011 CTGGAGCCGG TGAGCGTGGG TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT  
GACCTCGGCC ACTCGACCC AGAGCGCCAT AGTAACGTCG TGACCCCGGT CTACCATTGC GGAGGGCATA

=====  
amp  
=====

3081 CGTAGTTATC TACACGACGG GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT  
GCATCAATAG ATGTGCTGCC CCTCAGTCCG TTGATACCTA CTGTCTTAT CTGTCTAGCG ACTCTATCCA

*FIG. 6 (CONTINUED 3.)*

amp  
=====

3151 GCCTCACTGA TTAAGCATTG GTAACGTGCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAAC  
CGGAGTGACT AATTCGTAAC CATTGACAGT CTGGTTCAA TGAGTATATA TGAAATCTAA CTAAATTTTG

3221 TTCATTTTAA ATTTAAAAGG ATCTAGGTGA AGATCCTTTT TGATAATCTC ATGACCAAAA TCCCTTAACG  
AAGTAAAAAT TAAATTTTCC TAGATCCACT TCTAGGAAAA ACTATTAGAG TACTGGTTTT AGGGAATTGC

3291 TGAGTTTTCG TTCCACTGAG CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT  
ACTCAAAAGC AAGGTGACTC GCAGTCTGGG GCATCTTTTC TAGTTTCCTA GAAGAACTCT AGGAAAAAAA

3361 CTGCGCGTAA TCTGCTGCTT GCAACAAAAA AAACCACCGC TACCAGCGGT GGTGTTTTG CCGGATCAAG  
GACGCGCATT AGACGACGAA CGTTTGTTTT TTTGGTGGCG ATGGTCGCCA CCAAACAAAC GGCCTAGTTC

3431 AGCTACCAAC TCTTTTTCG AAGGTAACG GCTTCAGCAG AGCGCAGATA CCAAATACTG TCCTTCTAGT  
TCGATGGTTG AGAAAAAGGC TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTATTAGAC AGGAAGATCA

3501 GTAGCCGTAG TTAGGCCACC ACTTCAAGAA CTCTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG  
CATCGGCATC AATCCGGTGG TGAAGTTCTT GAGACATCGT GCGGATGTA TGGAGCGAGA CGATTAGGAC

3571 TTACCAGTGG CTGCTGCCAG TGGCGATAAG TCGTGCTTTA CCGGGTTGGA CTCAAGACGA TAGTTACCGG  
AATGGTCACC GACGACGGTC ACCGCTATTC AGCACAGAAT GGCCCAACCT GAGTTCTGCT ATCAATGGCC

3641 ATAAGGCGCA GCGGTCGGG TGAACGGGG GTTCGTGCAC ACAGCCCAGC TTGGAGCGAA CGACCTACAC  
TATTCCGCGT CGCCAGCCCG ACTTGCCCC CAAGCACGTG TGTCGGGTCG AACCTCGCTT GCTGGATGTC

3711 CGAACTGAGA TACCTACAGC GTGAGCATTG AGAAAGCGCC ACGCTTCCCG AAGGGAGAAA GCGGACAGG  
GCTTGACTCT ATGGATGTCG CACTCGTAAC TCTTTCGCGG TGCGAAGGGC TTCCCTCTTT CCGCCTGTCC

3781 TATCCCGTAA GCGGCAGGGT CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGAAAAC GCCTGGTATC  
ATAGGCCATT CGCCGTCCCA GCCTTGCTCT CTCGCGTGCT CCCTCGAAGG TCCCCCTTTG CGGACCATAG

3851 TTTATAGTCC TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTGG TGATGCTCGT CAGGGGGGCG  
AAATATCAGG ACAGCCCAA GCGGTGGAGA CTGAACTCGC AGCTAAAAAC ACTACGAGCA GTCCCCCGCG

3921 GAGCCTATGG AAAACGCCA GCAACGCGGC CTTTTTACGG TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC  
CTCGGATACC TTTTTCGGT CGTTGCGCCG GAAAAATGCC AAGGACCGGA AAACGACCGG AAAACGAGTG

3991 ATGTTCTTTC CTGCGTTATC CCCTGATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACCG  
TACAAGAAAG GACGCAATAG GGGACTAAGA CACCTATTGG CATAATGGCG GAACTCACT CGACTATGGC

4061 CTCGCCGCGAG CCGAACGACC GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCGCC CAATACGCA  
GAGCGGCGTC GGCTTGCTGG CTCGCGTGGC TCAGTCACTC GCTCCTTCGC CTTCTCGCGG GTTATGCGTT

4131 ACCGCCTCTC CCGCGCGGTT GGCGGATTCA TTAATGCAGC TGGCAGACA GGTTTCCCGA CTGGAAGCG  
TGGCGGAGAG GGGCGCGCAA CCGGCTAAGT AATTACGTCG ACCGTGCTGT CCAAAGGGCT GACCTTTCGC

4201 GGCAGTGAGC GCAACGCAAT TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTTA CACTTTATGC  
CCGTCACTCG CGTTGCGTTA ATTACACTCA ATCGAGTGAG TAATCCGTGG GGTCCGAAAT GTGAAATACG

4271 TTCCGGGTCG TATGTTGTGT GGAATTGTGA GCGGATAACA ATTTACACA GGAAACAGCT ATGACCATGA  
AAGGCCGAGC ATACAACACA CTTAACACT CGCCTATTGT TAAAGTGTGT CCTTTGTCGA TACTGGTACT

4341 TTACGCCA  
AATGCGGT

## SEQUENCE LISTING

&lt;110&gt; DEVGEN NV

&lt;120&gt; GENE EXPRESSION SYSTEM

&lt;130&gt; SCB/55177/001

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 5

&lt;170&gt; PatentIn Ver. 2.0

&lt;210&gt; 1

&lt;211&gt; 47

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
oligonucleotide o-GN59

&lt;400&gt; 1

ctagattaca actaattata cttatttgaa tattcaaatt ttcagac

47

&lt;210&gt; 2

&lt;211&gt; 47

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
oligonucleotide o-GN60

&lt;400&gt; 2

ccgggtctga aaatttgaa attcaaataa gtataattag ttgtaat

47

&lt;210&gt; 3

&lt;211&gt; 3498

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: plasmid  
pDW3123

&lt;400&gt; 3

agcttggcgc ctaatacgac tcactatagg gctgcaggtc gactctagat tacaactaat 60  
 tatacttatt tgaatattca aattttcaga cccgggattg gccaaaggac ccaaaggtat 120  
 gtttcgaatg atactaacat aacatagaac attttcagga ggacccttgg ctacgctcct 180  
 gctgggatta cacatggcat ggatgaacta tacaatagg gccggccgag ctccgcacgc 240  
 gccgctgtca tcagatcgcc atctcgcgcc cgtgcctctg acttctaagt ccaattactc 300  
 ttcaacatcc ctacatgctc tttctccctg tgctccacc ccctattttt gttattatca 360  
 aaaaaacttc ttcttaattt ctttgttttt tagcttcttt taagtcacct ctaacaatga 420  
 aattgtgtag attcaaaaat agaattaatt cgtaataaaa agtcgaaaaa aattgtgctc 480  
 cctcccccca ttaataataa ttctatccca aaatctacac aatgttctgt gtacacttct 540  
 tatgtttttt ttacttctga taaatttttt ttgaaacatc atagaaaaaa ccgcacacaa 600  
 aataccttat catatgttac gtttcagttt atgaccgcaa tttttatttc ttgcgacgtc 660  
 tgggcctctc atgacgtcaa atcatgctca tcgtgaaaaa gttttggagt atttttggaa 720

```

tttttcaatc aagtgaagtt ttatgaaatt aattttcctg cttttgcttt ttggggggttt 780
ccctatttgt ttgtcaagag ttctgaggac ggcgtttttc ttgctaaaaat cacaagtatt 840
gatgagcacg atgcaagaaa gatcggaaga aggttttgggt ttgaggctca gtggaaggtg 900
agtagaagtt gataatttga aagtggagta gtgtctatgg ggtttttgcc ttaaatgaca 960
gaatacattc ccaatatacc aaacataact gtttcctact agtcggccgt acggggccctt 1020
tcgtctcgcg cgtttcggtg atgacggtga aaacctctga cacatgcagc tcccgagac 1080
ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg gcgcgtcagc 1140
gggtgttggc ggggtgtcggg gctggcttaa ctatgcggca tcagagcaga ttgtactgag 1200
agtcacccat atgcggtgtg aaataccgca cagatgcgta aggagaaaat accgcatcag 1260
gcggccttaa gggcctcgtg atacgcctat ttttataggt taatgtcatg ataataatgg 1320
tttcttagac gtcaggtggc acttttcggg gaaatgtgcg cgaacccct atttgttat 1380
ttttctaaat acattcaaat atgtatccgc tcatgagaca ataacctga taaatgcttc 1440
aataatattg aaaaaggaag agtatgagta ttcaacattt ccgtgtcgcc ctattccct 1500
tttttgccgc attttgccct cctgtttttg ctacccaga aacgctggtg aaagtaaaag 1560
atgtcgaaga tcagttgggt gcacgagtg gttacatcga actggatctc aacagcggtta 1620
agatccttga gagttttcgc cccgaagaac gttttccaat gatgagcact tttaaagttc 1680
tgctatgtgg cgcggtatta tcccgattg acgcccggca agagcaactc ggtcgcgcga 1740
tacactattc tcagaatgac ttggttgagt actcaccagt cacagaaaag catcttacgg 1800
atggcatgac agtaagagaa ttatgcagt ctgccataac catgagtgat aacactgcgg 1860
ccaacttact tctgacaacg atcggaggac cgaaggagct aaccgctttt ttgcacaaca 1920
tgggggatca tgtaactcgc cttgatcgtt gggaaccgga gctgaatgaa gccataccaa 1980
acgacgagcg tgacaccacg atgcctgtag caatggcaac aacgttgcg aaactattaa 2040
ctggcgaact acttactcta gcttccgggc aacaattaat agactggatg gaggcggata 2100
aagttgcagg accacttctg cgctcggccc ttccggctgg ctggtttatt gctgataaat 2160
ctggagccgg tgagcgtggg tctcgggta tcattgcagc actggggcca gatggtaagc 2220
cctcccgat cgtagtattc tacacgacgg ggagtcaggc aactatggat gaacgaaata 2280
gacagatcgc tgagataggt gcctcactga ttaagcattg gtaactgtca gaccaagttt 2340
actcatatat actttagatt gatttaaaac ttcattttta atttaaaagg atctaggtga 2400
agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagttttcg ttccactgag 2460
cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tcctttttt ctgcgcgtaa 2520
tctgctgctt gcaaaacaaa aaaccaccgc taccagcggg ggtttgtttg ccggatcaag 2580
agctaccaac tctttttccg aaggttaactg gcttcagcag agcgcagata ccaaatactg 2640
tccttctagt gtagecgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat 2700
acctcgctct gctaactctg ttaccagtgg ctgctgccag tggcgataag tctgtcttta 2760
ccgggttggg ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg 2820
gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc 2880
gtgagcattg agaaagcgcc acgcttcccg aaggagagaa ggcggacagg tatccggtaa 2940
gcgcaggggt cggaacagga gagcgacga gggagcttcc aggggaaac gcctggtatc 3000
tttatagtc tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt 3060
cagggggggc gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg ttctggcct 3120
tttgcctggc ttttgctcac atgttcttct ctgcgttacc ccctgattct gtggataacc 3180
gtattaccgc ctttgagtga gctgataccg ctgcgcgag ccgaacgacc gagcgcagcg 3240
agtcagttag cgaggaagcg gaagagcgcc caatacgcaa accgcctctc cccgcgcgtt 3300
ggccgattca ttaatgcagc tggcacgaca gggttcccgta ctggaaagcg ggcagtgagc 3360
gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta cactttatgc 3420
ttccggtcgt tatgttgtgt ggaattgtga gcggataaca atttcacaca ggaacagct 3480
atgaccatga ttacgcca 3498

```

&lt;210&gt; 4

&lt;211&gt; 4348

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: plasmid  
pDW3124

&lt;400&gt; 4

```

agcttggcgc ctaatacgac tcaactatagg gctgcaggtc gactctagat tacaactaat 60
tatacttatt tgaatattca aattttcaga cccgggattg gccaaaggac ccaaagggtat 120
gtttcgaatg atactaacat aacatagaac aattttcagga ggacccttgg ctagcgtcga 180

```

cggtaccatg gggcgcgcca tgagtaaagg agaagaactt ttcactggag ttgtcccaat 240  
 tcttgttgaa ttagatgggtg atgttaatgg gcacaaatth tctgtcagtg gagagggtga 300  
 aggtgatgca acatacggaa aacttaccct taaattttatt tgcactactg gaaaactacc 360  
 tgttccatgg gtaagtttaa acatatatat actaactaac cctgattatt taaattttca 420  
 gccaacactt gtcactactt tctgttatgg tgttcaatgc ttctcgagat acccagatca 480  
 tatgaaacgg catgactttt tcaagagtgc catgcccgaa ggttatgtac aggaaagaac 540  
 tataattttt aaagatgacg ggaactacaa gacacgtaag tttaaacagt tcgggtactaa 600  
 ctaaccatac atattttaaat tttcaggtgc tgaagtcagg tttgagggtg atacccttgt 660  
 taatagaatc gagttaaaag gtatttgattt taaagaagat ggaaacattc ttggacacaa 720  
 attggaatac aactataact cacacaatgt atacatcatg gcagacaaac aaaagaatgg 780  
 aatcaaaagt gtaagtttaa acttggactt actaactaac ggattatatt taaattttca 840  
 gaacttcaaa attagacaca acattgaaga tgggaagcgtt caactagcag accattatca 900  
 acaaaatact ccaattggcg atggccctgt ccttttacca gacaaccatt acctgtccac 960  
 acaatctgcc ctctcgaaag atcccaacga aaagagagac cacatgggtc ttcttgagtt 1020  
 tgaacaagct gctgggatta cacatggcat ggatgaacta tacaataagg gccggccgag 1080  
 ctccgcacgc gccgctgtca tcagatcgcc atctcgcgcc cgtgcctctg acttctaagt 1140  
 ccaattactc ttaacatcc ctacatgctc tttctcctcg tgcctccacc cctatttttt 1200  
 gttattatca aaaaaacttc ttcttaattt ctttgttttt tagcttcttt taagtcacct 1260  
 ctaacaatga aattgttag attcaaaaat agaatttaatt cgttaataaaa agtcgaaaaa 1320  
 aattgtgctc cctcccccca ttaataataa ttctatccca aaatctacac aatgttctgt 1380  
 gtacacttct tatgtttttt ttacttctga taaatttttt ttgaaacatc atagaaaaaa 1440  
 ccgcacacaa aataccttat catatgttac gtttcagttt atgaccgcaa tttttatttc 1500  
 ttgcgcagtc tgggcctctc atgacgtcaa atcatgtcga tcgtgaaaaa gttttggagt 1560  
 attttttgaa tttttcaatc aagtgaagtt ttatgaaatt aattttcctg cttttgcttt 1620  
 ttgggggttt cccctattgt ttgtcaagag ttctcaggac ggctgttttt ttgctaaaaa 1680  
 cacaagtatt gatgagcacg atgcaagaaa gatcggaaga aggtttgggt ttgaggctca 1740  
 gtggaaggtg agtagaagtt gataatttga aagtggagta gtgtctatgg ggtttttgcc 1800  
 ttaaatgaca gaatacatc ccaatatacc aaacataact gtttcctact agtcggcctg 1860  
 acgggccctt tcgtctcgcg cgtttcggtg atgacgggtg aaacctctga cacatgcagc 1920  
 tcccgagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgctcagg 1980  
 gcgcgtcagc ggtgtgtggc ggtgttcggg gctggcctaa ctatgcggca tcagagcaga 2040  
 ttgtactgag agtcaccat atgcggtgtg aaataccgca cagatgcgta aggagaaaaa 2100  
 accgcataag gccgccttaa gggcctcgtg atacgcctat ttttataggt taatgtcatg 2160  
 ataataatgg ttcttagac gtcagggtgc acttttcggg gaaatgtgcg cggaacccct 2220  
 atttgtttat tttctaaat acattcaaat atgtatccgc tcatgagaca ataaccctga 2280  
 taaatgcttc aataatattg aaaaaggaag agtatgagta ttcaacattt ccgtgtcgcc 2340  
 cttattccct tttttgcggc attttgcctt cctgtttttg ctcacccaga aacgctgggt 2400  
 aaagtataag atgctgaaga tcagttgggt gcacgagtg gttacatcga actggatctc 2460  
 aacagcggta agatccttga gagttttcgc cccgaagaac gttttccaat gatgagcact 2520  
 ttaaaagttc tgctatgtgg cgcggtatta tcccgattg acgccgggca agagcaactc 2580  
 ggtcgccgca tacactatc tcagaatgac ttggttgagt actcaccagt cacagaaaaa 2640  
 catcttacgg atggcatgac agtaagagaa ttatgcagtg ctgccataac catgagtgat 2700  
 aacactgcgg ccaacttact tctgacaacg atcggaggac cgaaggagct aaccgctttt 2760  
 ttgcacaaca tgggggatca tgaactcgc cttgatcggt gggaaaccgga gctgaatgaa 2820  
 gccataccaa acgacgagcg tgacaccacg atgcctgtag caatggcaac aacgttgccg 2880  
 aaactattaa ctggcgaact acttactcta gcttcccgcc aacaattaat agactggatg 2940  
 gaggcggata aagtgcagg accacttctg cgtcgggcc ttccggctgg ctggtttatt 3000  
 gctgataaat ctggagccgg tgagcgtggg tctcgcggta tcattgcagc actggggcca 3060  
 gatggttaag cctcccgat cgtagtattc tacacgacgg ggagtcaggc aactatggat 3120  
 gaacgaaata gacagatcgc tgagataggt gcctcactga ttaagcattg gtaactgtca 3180  
 gaccaagttt actcatatat actttagatt gatttaaaac ttcattttta atttaaaagg 3240  
 atctaggtga agatcctttt tgataatctc atgaccaaaa tcccttaacg tgagttttcg 3300  
 ttccactgag cgtcagaccc cgtagaaaag atcaaaggat cttcttgaga tccctttttt 3360  
 ctgcgcgtaa tctgctgctt gcaacaaaaa aaaccaccgc taccagcggg ggtttgtttg 3420  
 ccggatcaag agctaccaac tctttttccg aaggttaactg gcttcagcag agcgagata 3480  
 ccaaatactg tcttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca 3540  
 ccgcctacat accctcgtct gctaactctg ttaccagtgg ctgctgccag tggcgataag 3600  
 tcgtgtctta ccgggttggg ctcaagacga tagttaccgg ataaggcgca gcggtcgggc 3660  
 tgaacggggg gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga 3720  
 tacctacagc gtgagcattg agaaagcgcc acgcttcccg aaggagagaa ggcggacagg 3780  
 tatccggtaa gcggcagggt cggaacagga gagcgacga gggagcttcc agggggaaac 3840

```

gcctgggtatc tttatagtc tgcggggttt cgccacctct gacttgagcg tcgatttttg 3900
tgatgctcgt caggggggag gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 3960
ttcctggcct tttgctggcc ttttgctcac atgttctttc ctgcgttacc ccctgattct 4020
gtggataacc gtattaccgc ctttgagtga gctgataccg ctgcgcgag ccgaacgacc 4080
gagcgagcag agtcagtgag cgaggaagcg gaagagcgcc caatacgcaa accgcctctc 4140
cccgcgcgtt ggccgattca ttaatgcagc tggcagcaca gggttcccga ctggaaagcg 4200
ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta 4260
cactttatgc ttccggctcg tatgttgtgt ggaattgtga gcggataaca atttcacaca 4320
ggaacagct atgacctga ttacgcca                                     4348

```

&lt;210&gt; 5

&lt;211&gt; 9309

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: plasmid pGN148

&lt;400&gt; 5

```

atgactgtc caaagaagaa gcgtaaggta ccgtaatga acacgattaa catcgctaag 60
aacgacttct ctgacatcga actggctgct atcccggtca acactctggc tgaccattac 120
ggtgagcgtt tagctcggtg agtttaaaca tctagatact aactaacgat taacatttaa 180
attttcagcg aacagttggc ccttgagcat gactcttacg agatgggtga agcacgcttc 240
cgcaagatgt ttgagcgtca acttaaagct ggtgaggttg cggataacgc tggcccaaag 300
cctctcatca ctaccctact ccctaagatg attgcacgca tcaacgactg gtttgaggaa 360
gtgaaagcta agcgcgcaa gcgcccga caactccagt tcctgcaaga aatcaagccg 420
gaagccgtag cgtacatcac cattaagacc actctggctt gcctaaccag tgctgacaat 480
acaaccgttc aggtgtagc aagcgcaatc ggtcgggcca ttgaggacga ggctcgcttc 540
ggtcgtatcc gtgacctga agctaagcac ttcaagaaaa acgttgagga acaactcaac 600
aagcgctgag ggacgtcta caagaaagca tttatgcaag ttgtcgaggg tgacatgtct 660
tctaagggtc tactcggtgg cgagggcgtg tcttcgtggc ataaggaaga ctctattcat 720
gtaggagtac gctgcatcga gatgtcatt gactcaaccg gactgggttag cttacaccgc 780
caaaatgctg cggtagtagg tcaagactct gagactatcg aactcgaccc tgaatagcct 840
gaggtatcgc caaccgtgc aggtgcgctg gctggcatct ctccgatgtt ccaaccttgc 900
gtagttcttc ctaagccgtg gactggcatt actggtgggt gctattgggc taacggtcgt 960
cgctctctgg cgctgggtcg tactcacagt aagaagcac tgatgcgcta cgaagacgtt 1020
tacctgctg aggtgtacaa agcgattaac attcgcaaaa acaccgcatg gaaaatcaac 1080
aagaaagtcc tagcggtcgc caacgtaatc accaagtggg agcattgtcc ggtcgaggac 1140
atccctgcga ttgagcgtga agaactcccg atgaaaccgg aagacatcga catgaatcct 1200
gaggtcttca ccgctgggaa acgtgctgcc gctgctgtgt accgcaagga caaggtcgc 1260
aagtctcgcc gtatcagcct tgagttcatg cttgagcaag ccaataagtt tgctaaccat 1320
aaggccatct ggttccctta caacatggac tggcgcggtc gtgtttacgc tgtgtcaatg 1380
ttcaaccgcg aagctaacga tatgaccaa ggactgctta cgctggcgaa aggtaaacca 1440
atcggttaag aaggttacta ctggctgaaa atccacggtg caaactgtgc ggggtgcgat 1500
aaggttccgt tcctgagcg catcaagttc attgaggaag accacgagaa catcatggct 1560
tgcgctaagt tccactgga gaacacttgg tgggctgagc aagattctcc gttctgcttc 1620
cttgcttct gctttgagta cgctggggtg cagcaccacg gcctgagcta taactgctcc 1680
cttccgctgg cgtttgacgg gtcttgctct ggcatccagc acttctccgc gatgctccga 1740
gatgaggtag gtggtcgcg ggttgtaagt ttaactcta tcctactaac taacgaagct 1800
tatttaaatt ttcagaactt gcttcctagt gaaaccgttc aggacatcta cgggattgtt 1860
gctaagaaag tcaacgagat tctacaagca gacgcaatca atgggaccga taacgaagta 1920
gttaccgtga ccgatgagaa cactggtgaa atctctgaga aagtcaagct gggcactaag 1980
gcactggctg gtcaatggct ggcttacggt gttactcgca gtgtgactaa gcgttcagtc 2040
atgacgctgg ctacgggtc caaagagttc ggcttccgtc aacaagtgct ggaagatacc 2100
attcagccag ctattgatcc cggcaagggt ctgatgttca ctacgcccga tcaggctgct 2160
ggatacatg ctaagctgat ttgggaatct gtgagcgtga cggtggtagc tgcggttgaa 2220
gcaatgaact ggcttaagtc tgctgctaag ctgctggctg ctgaggtcaa agataagaag 2280
actggagaga ttcttcgcaa gcgttgcgct gtgcattggg tcaactccga tggtttcct 2340
gtgtggcagg aatacaagaa gcctattcaa acgcttttga acctgatgtt cctcggtcag 2400
ttccgcttac agcctaccat taacaccaac aaagatagcg agattgatgc acacaaacag 2460
gagtctggtg tcgctcctaa cttgttacac agccaagacg gtgaccacct tcgtaagact 2520

```



gtagtgtggg	cacacgagaa	gtacggaatc	gaatcttttg	caactgattca	cgactccttc	2580
ggtaccattc	cggtgacgc	tgcgaacctg	ttcaaagcag	tgcgcgaaac	tatggttgac	2640
acatatgagt	cttgtgatgt	actggctgat	ttctacgacc	agttcgctga	ccagttgcac	2700
gagttctaat	tggacaaaat	gccagcactt	ccggtctaaag	gtaacttgaa	cctccgtgac	2760
atcttagagt	cggacttcgc	gttcgcgtaa	gaattccaac	tgagcgccgg	tcgtaccat	2820
taccaacttg	tctggtgtca	aaaataatag	gggcgcgtgt	catcagagta	agtttaaaact	2880
gagttctact	aactaacgag	taatatattaa	attttcagca	tctcgcgccc	gtgcctctga	2940
cttctaagtc	caattactct	tcaacatccc	tacatgctct	ttctccctgt	gtccccacc	3000
cctatttttg	ttattatcaa	aaaaacttct	tcttaatttc	tttgtttttt	agcttctttt	3060
aagtccacct	taacaatgaa	attgtgtaga	ttcaaaaata	gaattaattc	gtaataaaaa	3120
gtcgaaaaaa	atttgtctcc	ctcccccat	taataataat	tctatcccaa	aactctacaca	3180
atgttctgtg	tacacttctt	atgttttttt	tacttctgat	aaattttttt	tgaacatca	3240
tagaaaaaac	cgcacacaaa	atacccttct	atagtgtacg	tttcagttta	tgaccgcaat	3300
ttttatttct	tcgcacgtct	gggcctctca	tgacgtcaaa	tcatgctcat	cgtgaaaaag	3360
ttttggagta	tttttggaa	ttttcaatca	agtgaagt	tatgaatta	attttctctg	3420
ttttgctttt	tgggggtttc	ccctattgtt	tgcaagagt	ttcgaggacg	gcgtttttct	3480
tgctaaaatc	acaagtattg	atgagcacga	tgcaagaaag	atcggaagaa	ggtttgggtt	3540
tgaggctcag	tggaaagtga	gtagaagtgt	ataatttgaa	agtggagtag	tgcttatggg	3600
gtttttgcct	taaagtacag	aatacattcc	caatatacca	aacataactg	tttctacta	3660
gtcggccgta	cgggcccttt	cgtctcgcg	gttcggtga	tgacggtgaa	aacctctgac	3720
acatgcagct	cccgagacg	gtcacagctt	gtctgtaagc	ggatgccggg	agcagacaag	3780
cccgtcaggg	cgcgtcagcg	ggtgttgccg	ggtgtcgggg	ctggcttaac	tatgcggcat	3840
cagagcagat	tgtactgaga	gtgcaccata	tgcggtgtga	aataccgcac	agatgcgtaa	3900
ggagaaaaa	ccgcacagc	cggccttaag	ggcctcgtga	tacgcctatt	tttataggtt	3960
aatgtcatga	taataatggt	ttcttagacg	tcagggtgca	cttttcgggg	aaatgtgcgc	4020
ggaaccccta	tttgtttatt	tttctaaata	cattcaaata	tgtatccgct	catgagacaa	4080
taacctgat	aatgcttca	ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	4140
cgtgtcgccc	ttattccctt	ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	4200
acgctggtga	aagttaaaga	tgctgaagat	cagttgggtg	cacgagtggt	ttacatcgaa	4260
ctggatctca	acagcgttaa	gatccttgag	agttttcgcc	ccgaagaacg	ttttccaatg	4320
atgagcactt	ttaaagttct	gctatgtggc	gcggtattat	cccgtattga	cgccgggcaa	4380
gagcaactcg	gtcgcgcgat	acactattct	cagaatgact	tggttgagta	ctcaccagtc	4440
acagaaaagc	atcttacgga	tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	4500
atgagtata	acactgcggc	caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	4560
accgcttttt	tgcaacaacat	gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	4620
ctgaatgaag	ccataccaaa	cgcagagcgt	gacaccacga	tgctgtagc	aatggcaaca	4680
acgttgcgca	aactattaac	tggcgaaata	cttactctag	cttcccgcca	acaattaata	4740
gactggatgg	aggcggataa	agtgcagga	ccacttctgc	gctcgccct	tcgggctggc	4800
tggtttattg	ctgataaatc	tggagccggt	gagcgtgggt	ctcgcggtat	cattgcagca	4860
ctggggccag	atggttaagc	ctcccgatc	gtagtattct	acacgacggg	gagtcaggca	4920
actatggatg	aacgaaatag	acagatcgct	gagataggtg	cctcactgat	taagcattgg	4980
taactgtcag	accaagttta	ctcatatata	ctttagattg	atttaaaact	tcatttttaa	5040
tttaaaagga	tctagtgtaa	gatccttttt	gataatctca	tgacaaaaat	cccttaacgt	5100
gagttttcgt	tccactgagc	gtcagacccc	gtagaaaaga	tcaaaggatc	ttcttgagat	5160
cctttttttc	tgcgcgtaat	ctgctgcttg	caaacaaaaa	aaccaccgct	accagcgggtg	5220
gtttgtttgc	cggatcaaga	gtaccaact	ctttttccga	aggtaactgg	cttcagcaga	5280
gcgcagatac	caaatactgt	ccttctagt	tagccgtagt	taggccacca	cttcaagaac	5340
tctgtagcac	cgcctacata	cctcgctctg	ctaatacctgt	taccagtggc	tgctgccagt	5400
ggcgataagt	cgtgtcttac	cgggttggtg	tcaagacgat	agttaccgga	taaggcgag	5460
cggtcgggct	gaacgggggg	ttcgtgcaca	cagcccagct	tgagcgcaac	gacctacacc	5520
gaactgagat	acctacagcg	tgagcattga	gaaagcgcca	cgttcccgga	agggagaaaag	5580
gcggacaggt	atccggtaa	cggcagggtg	ggaacaggag	agcgcacgag	ggagcttcca	5640
gggggaaacg	cctggatatc	ttatagtcct	gtcgggtttc	gccacctctg	acttgagcgt	5700
cgatttttgt	gatgctcgtc	aggggggcgg	agcctatgga	aaaacgccag	caacgcggcc	5760
tttttacggg	tcttgccctt	ttgctggcct	ttgtctcaca	tgttctttcc	tcggttatcc	5820
cctgattctg	tggataaccg	tattaccgcc	tttgagttag	ctgataccgc	tcgccgcagc	5880
cgaacgacgg	agcgcagcga	gtcagtgagc	gaggaagcgg	aagagcgccc	aatacgcaaa	5940
ccgcctctcc	ccgcgcgttg	gcgattcat	taatgcagct	ggcacgcagc	gtttcccgac	6000
tggaaagcgg	gcagtgagcg	caacgcaatt	aatgtgagtt	agctcactca	ttaggcaccc	6060
caggctttac	actttatgct	tccggctcgt	atgttggtg	gaattgtgag	cggataacaa	6120
tttcacacag	gaaacagcta	tgacctgat	tacgccaaag	tgtaagttta	aacatgatct	6180

```

tactaactaa ctattctcat ttaaattttc agagcttaaa aatggctgaa atcactcaca 6240
acgatggata cgctaacaac ttggaaatga aataagcttg catgcctgca gagcaaaaaa 6300
atactgcttt tccttgcaaa attcgggtgct ttcttcaaag agaaactttt gaagtcggcg 6360
cgagcatttc cttctttgac ttctctcttt ccgccaaaaa gcctagcatt tttattgata 6420
atttgattac acacactcag agttcttcga catgataaag tgtttcattg gcaactgccc 6480
taacagtaca tgacaagggc ggattattat cgatcgatat tgaagacaaa ctccaaatgt 6540
gtgctcattt tggagcccg tgtggggcag ctgctctcaa tataattacta gggagacgag 6600
gagggggacc ttatcgaacg tcgcatgagc cattctttct tctttatgca ctctcttcac 6660
tctctcacac attaatcgat tcatagactc ccatattcct tgatgaagggt gtgggttttt 6720
agctttttt cccgatttgt aaaaggaaga ggctgacgat gttaggaaaa agagaacgga 6780
gccgaaaaaa aattaaaaga atttgcctct tttaagccga cactttttag acagcattcg 6840
ccgctagtgt tgaagtttaa attttaaaaa ataaaaatta gtttcaattt tttttaatta 6900
ctaaataggc aaaagttttt tcaagaactc tagaaaaact agcttaattc atgggtacta 6960
gaaaaattct tgttttaaat ttaatattta tcttaagatg taattacgag aagctttttt 7020
gaaaattctc aattaaaaga atttgccgat ttagaataaa agtcttcaga aatgagtaaa 7080
agctcaaat agaagtttgt ttttaagga aaaacacgaa aaaagaacac tatttatctt 7140
ttctccccg cgtaaaatta gttgttgtga taatagtgat ccgctgtcta tttgcaactc 7200
gctcttcaca ccgtgcttcc tctcacttga cccaacagga aaaaaaaca tcacgtctga 7260
gacgggtgat tgccttatca agagcgtcgt ctctttcacc cagtaacaaa aaaaatttgg 7320
tttctttact ttatatattat gtaggtcaca aaaaaaaagt gatgcagttt tgtgggtcgg 7380
ttgtctccac accacctccg cctccagcag cacacaatca tcttcgtgtg ttctcgacga 7440
ttcttgttat gcccggtgct tgaatgcacc acattcgacg cgcaactaca caccacactc 7500
actttcgggtg gtattactac acgtcatcgt tgttcgtagt ctcccgctct ttctgcccga 7560
ctcactcttc attattcccc ttggtgtatt gatttttttt aaatggtaca ccactcctga 7620
cgtttctacc ttcttgtttt ccgtccattt agattttatc tggaaaattt tttaaaattt 7680
taggccagag agttctagtt cttgttctaa aagcttaggt cagacataca ttttctattt 7740
ctcatcaaaa aaaaagttga taaagaaaac tggttattca gaaagagtgt gtctcgttga 7800
aattgattca aaaaaaatt cccaccctc gcttgtttct caaaatatga gatcaacgga 7860
ttttttcctt ctcgattcaa ttttttgcgt cgctctgtct gccaaagtgt gtgtgtccga 7920
gcaaagatg agagaattta caaacagaaa tgaaaaaaag ttggccaaat aatgaagttt 7980
tatccgagat tgatgggaaa gatatttaatt ttctttacgg tttggagggg agagagagat 8040
agattttcgc atcaaaactcc gcctttttaca tgtcttttag aatctaaaaat agatttttct 8100
catcattttt aatagaaaaat cgagaaaatta cagtaatttc gcaattttct tgccaaaaat 8160
acacgaaatt tgtgggtctc gccacgatct cggtcttagt ggttcatttg gtttaaaagt 8220
ttataaaatt tcaaatctta gtgtttaatt tccgcataat tggacctaaa atgggttttt 8280
gtcatcattt tcaacaagaa atcgtgaaaa tcctgttggt tdcgcaattt cttttcaaaa 8340
atacacgaaa tatatggtaa tttcccgaat tattgagggt ctgccacga tttcagtcac 8400
agtggccag agttatcacg aaaaaagttc gcctagtctc acattttccg aaaaaccgaat 8460
ctaaattagt tttttgtcat cattttgaac aaaaaatcga gacatcccta tagtttcgca 8520
attttcgtcg cttttctctc caaaaatgac agtctagaat taaaattcgc tggaaactgg 8580
accatgatat cttttctccc cgtttttcat tttatttttt attacactgg attgactaaa 8640
ggtcaccacc accgccagtg tgtgccatat cacacacaca cacacacaca atgtcgagat 8700
tttatgtgtt atccctgctt gatttcgttc cgttgtctct ctctctctat tcatcttttg 8760
agccgagaag ctccagagaa tggagcacac aggatcccgg cgcgcgatgt cgtcgggaga 8820
tggcgccgcc tgggaagccg ccgagagata tcagggaaga tcgtctgatt tctcctcgga 8880
tgccacctca tctctcagat ttctccgcct gttactccct gccgaacctg atatttccc 8940
ttgtcgtaaa gagatgtttt tattttactt tacacgggt cctctctctc tgccagcaca 9000
gctcagtggt ggctgtgtgc tcgggtcct gccaccggcg gcctcatctt cttctcttc 9060
ttctctcctg ctctcgctta tcaactcttc attcattctt attccttttc atcatcaaac 9120
tagcatttct tactttattt atttttttca attttcaatt ttcagataaa accaaactac 9180
ttgggttaca gccgtcaaca gatccccggg attggccaaa ggaccctaaag gtatgtttcg 9240
aatgatacta acataacata gaacattttc aggaggacc ttgcttgag ggtaccgagc 9300
tcagaaaaa 9309

```